University of Alberta

The Construction and Evaluation of a Multiple Capillary DNA Sequencer

by

Sue J. Bay

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Chemistry

Edmonton, Alberta

Fall, 1998

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-34733-8

Canada

University of Alberta

Library Release Form

Name of Author: Sue Jee Bay

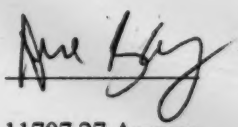
Title of Thesis: The Construction and Evaluation of a Multiple Capillary DNA Sequencer

Degree: Doctor of Philosophy

Year this Degree Granted: 1998

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever with the author's written permission.



11707 27 Avenue
Edmonton, Alberta
T6J 3N7

June 22, 1998

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **The Construction and Evaluation of a Multiple Capillary DNA Sequencer** submitted by Sue Bay in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Dr. N.J. Dovichi

Dr. L. Li

Dr. G. Kotovych

Dr. W.E. Harris

Dr. J.F. Elliott

Dr. T.K. Christopoulos

4 May 1998

ABSTRACT

DNA sequencing has medical applications because it may identify genes responsible for human disease and the pathogens that cause infectious disease. Capillary gel electrophoresis DNA sequencers are faster and require less sample than current slab gel instruments. The simultaneous analysis of many samples and automated sample injection are the primary advantages of capillary sequencers.

A 16-capillary DNA sequencer was built with a micromachined sheath-flow cuvette for post-column, laser-induced fluorescence detection. Photolithography techniques etched micrometer structures in the cuvette to precisely position the capillaries. The glass structures aligned the fluorescent signals from each capillary to the avalanche photodiode (APD) detectors. The first construction of the sequencer suffered from poor sensitivity because of the noise in the locally constructed APD circuit. The APD circuit was rebuilt and signal-shot-noise-limited. The 16-capillary instrument detection limit (3σ) was 5×10^{-12} M (10 000 molecules) of fluorescein.

A diagnostic to identify pathogenic bacteria by DNA sequencing was developed in collaboration with an infectious diseases group. A single set of Polymerase Chain Reaction (PCR) primers amplified a portion of the Heat Shock Protein-60 kDa (HSP60) gene in several *Staphylococcus* species. DNA sequence variations in the single copy and highly conserved gene may identify bacterial species and strains. The 16-capillary instrument sequenced the HSP60 PCR products from six reference *Staphylococcus* species (*S. aureus* 8325-4, *S. epidermidis* 9759, *S. haemolyticus* ATCC 29970, *S. lugdunensis* CRSN 850412, *S. saprophyticus* KL122

and, *S. schleiferi* ATCC 43808). The 15% sequence dissimilarity of the six *Staphylococcus* species supports HSP60 DNA sequencing for species identification.

The accuracy of the DNA sequencing base-calling software was evaluated. BASS, a public domain base-calling program, was locally modified for capillary sequencing. The processed data required significant manual editing which improved sequencing accuracy by approximately 4%. An average read lengths of the 16-capillary sequencer were about 500 bases at >97% in 3 h.

The 16-capillary instrument can sequence at least 500 bases of finished sequence in 3 h. Assuming 4-hour turnover time, 3 runs per day and 5 runs per week, the instrument can sequence at a rate of 120 kilobases per week.

To my parents

ACKNOWLEDGEMENTS

I would like to thank my supervisor Norm Dovichi for his guidance and insight. John Crabtree, my co-conspirator, developed the photolithography procedures for the micromachined cuvette and equally contributed to the first-generation version of the 16-capillary sequencer. Darren Lewis was an incredible resource; he constructed the data-acquisition hardware and software and greatly contributed to rebuilding the instrument. This thesis could not have been written without their efforts.

There are many other people I must thank for their contributions: Jianzhong Zhang for his multiple capillary instrumentation experience; Larry Coulson for his APD design and noise isolation advice; Xiaoling Puyang for her cloning expertise; Glen Fitzpatrick, Tran Tran, and Graham McKinnon at the Alberta Microelectronics Centre for use of their photolithography facilities; Stephen Lewis, who modified BASS, the base-calling software; and John Bercovitz at Lawrence Berkeley National Labs for building the gel pressure refiller. I also thank Dr. Anthony Chow at University of British Columbia for his support and collaboration in the *Staphylococcus* sequencing project. I would also like to recognize the work of Pieter Roos, Sarah Kwong and Dawn Richards, the University of Alberta Chemistry Machine and Electronic Shops, Kym Schreiner and Costas Stathakis for their editing comments, and everyone at the Northern Lights Laser Lab.

My funding for graduate school came from scholarships from University of Alberta, the Natural Sciences and Engineering Research Council of Canada, and the Alberta Heritage Foundation for Medical Research.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

1.1. INTRODUCTION 2

1.2. DNA SEQUENCING 3

1.3. MULTIPLE-CAPILLARY INSTRUMENTS 7

1.4. *STAPHYLOCOCCUS* IDENTIFICATION 11

1.5. THIS THESIS 14

1.6. REFERENCES 15

CHAPTER 2: THE 16-CAPILLARY SEQUENCER, VERSION 1.0 20

2.1. INTRODUCTION 21

2.1.1. Separation Apparatus 21

Glass Micromachining 21

2.1.2. Detection Apparatus 25

Optics 25

Avalanche Photodiodes 27

2.1.3. DNA Sequencing 32

2.2. EXPERIMENTAL 32

2.2.1. Separation Apparatus 32

2.2.1.1. Micromachined Sheath-flow Cuvette 32

2.2.1.2. Plexiglas Heating Box and Sample Injector 35

2.2.2. Detector Apparatus 39

2.2.2.1. Optics 39

2.2.2.2. Avalanche Photodiode Cooling 40
2.2.2.3. Avalanche Photodiode Electronic Circuitry 44
2.2.2.4. Avalanche Photodiode Biasing Voltage 46
2.2.2.5. Chopper/Filter Wheel and Data Acquisition (DAQ) 46
2.2.3. DNA Sequencing 48
2.2.3.1. Single-Stranded Template Sequencing 48
2.2.3.2. Cloned Wild-Type M13 Sequencing 50
2.3. RESULTS AND DISCUSSION 51
2.3.1. Separation Apparatus 52
2.3.1.1. Micromachined Sheath-flow Cuvette 52
2.3.1.2. Plexiglas Heating Box and Sample Injector 59
2.3.2. Detector Apparatus 60
2.3.2.1. Optics 60
2.3.2.2. Avalanche Photodiode Cooling 60
2.3.2.3. Avalanche Photodiode Electronic Circuitry 63
2.3.2.4. Avalanche Photodiode Biasing Voltage 65
2.3.2.5. Chopper/Filter Wheel and Data Acquisition 66
2.3.3. DNA Sequencing 67
2.3.3.1. Single-Stranded Template Sequencing 67
2.3.3.2. Cloned Wild-Type M13 Sequencing 67
2.4. CONCLUSIONS 71
2.5. REFERENCES 72

CHAPTER 3: 16-CAPILLARY SEQUENCER, VERSION 2.0 74

3.1. INTRODUCTION 75

3.1.1. Capillary Gel Refiller 75

3.2. EXPERIMENTAL 76

3.2.1. Separation Apparatus 76

3.2.1.1. Capillary Gel Refiller 76

3.2.1.2. Sample Injector 82

3.2.1.3. Peltier Capillary Heater 88

3.2.2. Detector Apparatus 88

3.2.2.1. Optics 88

3.2.2.2. Avalanche Photodiode Cooling 92

3.2.2.3. Avalanche Photodiode Electronic Circuitry 92

3.2.2.4. Avalanche Photodiode Biasing Voltage 95

3.2.3. Capillary Electrophoresis 95

3.2.3.1. Gel Preparation 95

3.2.3.2. DNA Sequencing Samples 97

3.2.3.3. Sample Injection and Electrophoresis 99

3.3. RESULTS AND DISCUSSION 100

3.3.1. Separation Apparatus 100

3.3.1.1. Capillary Gel Refiller 100

3.3.1.2. Sample Injector 104

3.3.1.3. Peltier Capillary Heater 105

3.3.2. Detection Apparatus....106
3.3.2.1. Optics....106
3.3.2.2. Avalanche Photodiode Cooling....107
3.3.2.3. Avalanche Photodiode Electronic Circuitry....110
3.3.2.4. Avalanche Photodiode Biasing Voltage....110
3.4. CONCLUSIONS....111
3.5. REFERENCES....112
CHAPTER 4: OPTIMIZING THE AVALANCHE PHOTODIODES....113
4.1. INTRODUCTION....114
4.1.1. Noise....114
4.1.2. Avalanche Photodiodes....116
4.1.3. Limit of Detection....117
4.2. EXPERIMENTAL....118
4.2.1. Noise....118
4.2.1.1. Noise Intensity *vs.* Frequency at Increasing APD Bias....118
4.2.1.2. Signal-to-Noise Studies with Increasing APD Bias and External Gain....118
4.2.2. Limit of Detection....120
4.2.2.1. Fiber Optic Core Comparison on 5-Capillary Instrument....120
4.2.2.2. LOD Measurements on 16-Capillary Instrument....120
4.3. RESULTS AND DISCUSSION....121
4.3.1. Noise....121
4.3.1.1. Noise Intensity *vs.* Frequency at Increasing APD Bias....121

4.3.1.2. Signal-to-Noise Studies with Increasing APD Bias and External Gain.124
4.3.1.3. Characterization of the Noise in APDs.....132
4.3.1.4. Calculation of Noise in APD.....132
4.3.2. Limit of Detection.....137
4.3.2.1. Fiber Optic Core Comparison on 5-Capillary Instrument.....137
4.3.2.2. LOD Measurements on 16-Capillary Instrument.....139
4.4. CONCLUSIONS.....143
4.5. REFERENCES.....144
CHAPTER 5: *STAPHYLOCOCCUS* HSP60 GENE SEQUENCING.....145
5.1. INTRODUCTION.....146
5.1.1. Phylogenetic Tree Construction.....146
5.2. EXPERIMENTAL.....153
5.2.1. Bacterial Isolates.....153
5.2.2. HSP60 PCR and Cloning.....153
5.2.2.1. Boiling Lysis HSP60 PCR.....153
5.2.2.2. Cloning the HSP60 PCR Products into pUC19.....156
5.2.2.3. Boiling/Lysis AB PCR.....158
5.2.3. Subcloning the Halved HSP60 PCR Products.....160
5.2.3.1. AB PCR.....160
5.2.3.2. *Eco*RI/*Bam*HI Digest.....160
5.2.3.3. Halving Digest.....163
5.2.3.4. Cloning the Halved HSP60 Gene Products into pUC19.....163

5.2.4. DNA Sequencing 164
5.2.5. Gel Preparation 165
5.2.6. Sample Injection and Run 166
5.2.7. DNA Sequence Data Processing 166
5.2.8. Phylogenetic Tree Construction 167
5.2.9. Nucleotide Accession Numbers 167
5.3. RESULTS AND DISCUSSION 168
5.3.1. HSP60 Gene Sequences 168
5.3.1.1. Phylogenetic Tree 168
5.3.1.2. Cloning HSP60 PCR Products 176
5.3.2. Capillary DNA Sequencing 178
5.4. CONCLUSIONS 181
5.5. REFERENCES 181
CHAPTER 6: ACCURACY OF CAPILLARY SEQUENCING 185
6.1. INTRODUCTION 186
6.1.1. Background 186
6.2. EXPERIMENTAL 187
6.2.1. DNA Sequencing Data 187
6.2.2. Data Processing and Editing 187
6.2.3. Error Analysis 190
6.2.3.1. Data Categories 190
6.2.3.2. Accuracy *vs.* Base Position 191

6.2.3.3. Distribution of Errors *vs.* 50-Base Windows 191

6.2.3.4. Read Lengths 191

6.3. RESULTS AND DISCUSSION 191

6.3.1. Accuracy *vs.* Base Position 191

6.3.2. Distribution of Errors vs. 50-Base Windows 194

6.3.3. Read Lengths 209

6.4. CONCLUSIONS 211

6.5. REFERENCES 212

CHAPTER 7: DNA SEQUENCING ARTIFACTS 213

7.1. INTRODUCTION 214

7.2. EXPERIMENTAL 214

7.2.1. DNA Sequencing Samples 214

7.2.2. Gel Preparation 217

7.2.3. Sample Injection and Run 217

7.3. RESULTS AND DISCUSSION 218

7.4. CONCLUSIONS 225

7.5. REFERENCES 225

CHAPTER 8: CONCLUSIONS AND FUTURE WORK 226

8.1. CONCLUSIONS 227

8.2. REFERENCES 229

APPENDIX I: DNA SEQUENCING REAGENTS 231

APPENDIX II: PARTIAL HSP60 GENE SEQUENCES 233

LIST OF FIGURES

Figure 1.1. The Structure of DNA. 4

Figure 1.2. Base-Pairing of DNA....5

Figure 1.3. Sanger DNA Sequencing....6

Figure 2.1. The Separation Apparatus. 22

Figure 2.2. The Detector Apparatus....23

Figure 2.3. Basic Steps of Photolithography....24

Figure 2.4. Capillary Alignment in the Micromachined Cuvette....26

Figure 2.5. Conventional and GRIN Lenses. 28

Figure 2.6. Diagram of a Reach Through Structure APD....30

Figure 2.7. Current-Voltage Curve for a Silicon Avalanche Photodiode. 31

Figure 2.8. Micromachined Cuvette Assembly....34

Figure 2.9. The 16-Capillary Cuvette Mounted in Holder....36

Figure 2.10. Plexiglas Capillary Box with High Voltage and Forced Air Heating....38

Figure 2.11. The GRIN Lens Image in the Cuvette. 41

Figure 2.12. Styrofoam APD Box....43

Figure 2.13. The Avalanche Photodiode Electronics....45

Figure 2.14. The LED Coupling to FC Connector....47

Figure 2.15. The 16-Capillary DNA Sequencer....53

Figure 2.16. The First Generation 16-Capillary Sequencer. 55

Figure 2.17. The Fluorescent Spots in the Micromachined Cuvette....57

Figure 2.18. The Arrangement of the GRIN Lenses. 61

Figure 2.19. Input Connections for Transimpedance Amplifiers. 64

Figure 2.20. A 16-Capillary Run of Single-Stranded M13mp18. 68

Figure 2.21. Wild-Type M13 Sequencing Run. 69

Figure 3.1. The Capillary Gel Refiller. 77

Figure 3.2. Top View of the 16-Capillary Gel Refiller. 78

Figure 3.3. The 16-Capillary Gel Refiller. 80

Figure 3.4. The Sample Injection Apparatus. 83

Figure 3.5. The Sample Injector. 84

Figure 3.6. The 96-Well Gold-Plated Microtitre Plate. 86

Figure 3.7. Peltier Capillary Heater. 89

Figure 3.8. The Peltier Capillary Heater. 90

Figure 3.9. APD Cooling Box 93

Figure 3.10. APD Electronic Circuitry. 94

Figure 3.11. DNA Sequencing with Capillary Refilled with DMA at Room Temp 102

Figure 3.12. DNA Sequencing at 40°C with Capillary Refilled at Room Temperature. ..103

Figure 3.13. Detection of Sample A with Two Different Optical Filters Sets 108

Figure 3.14. Detection of Sample B with Two Different Optical Filters Sets 109

Figure 4.1. The LED in the Signal-to-Noise Experiments 119

Figure 4.2. Dark Current vs. Frequency of APD Output at Increasing Bias 122

Figure 4.3. Dark Current vs. Frequency of APD Output at Increasing Bias 123

Figure 4.4. APD Response vs. APD Bias at External Gain of 10^9 V/A 125

Figure 4.5. APD Response to LED Light Intensity at APD External Gain 10^9 V/A 126

Figure 4.6. APD Response to LED Light Intensity at APD External Gain 10^{10} V/A.....127

Figure 4.7. Log Signal-to-Noise vs. Log Intensity at External Gain of 10^{10} V/A.....129

Figure 4.8. Signal-to-Noise Ratio vs. APD Bias at External Gain of 10^{10} V/A 130

Figure 4.9. Log (S/N) vs. Log (S) at External Gain of 10^{9} V/A 133

Figure 4.10. Log (S/N) vs. Log (S) at External Gain of 10^{10} V/A 134

Figure 4.11. Log (S/N) vs. Log (S) of Single Photon Counting Module.....135

Figure 4.12. Intrinsic Factors of Fiber Optic Connector Loss 140

Figure 4.13. Injection of 2×10^{-11} M Fluorescein.....141

Figure 4.14. Raw Data of the Injection of 2×10^{-11} M Fluorescein.....142

Figure 5.1. Phylogenetic Trees.....148

Figure 5.2. HSP60 PCR and cloning 154

Figure 5.3. The HSP60 PCR of 6 *Staphylococcus* Species.....157

Figure 5.4. The pUC19 Plasmid Multiple Cloning Site.....159

Figure 5.5. Subcloning the Halved HSP60 PCR Products.....161

Figure 5.6. Unrooted Phylogenetic Tree of *Staphylococcus* HSP60 Gene Sequences ...154

Figure 5.7. Unrooted Phylogenetic Tree of *Staphylococcus* HSP60 Gene Sequences and other Closely Related Organisms from BLAST Search.....154

Figure 5.8. DNA Sequence Alignment of *S. aureus* 8325-4 and *S. aureus* 912.....174

Figure 5.9. Amino Acid Sequence Alignment of *S. aureus* 8325-4 and *S. aureus* 912..175

Figure 5.10. HSP60 PCR with Modified PCR Primers 177

Figure 5.11. The HSP60 Gene Sequence of *S. epidermidis*.....179

Figure 6.1. The Mobility Shift of G-Terminated (TAMRA) Fragments 188

Figure 6.2. The Mobility Shift of T-Terminated (ROX) Fragments 189

Figure 6.3. Accuracy vs. Base Position of All Forward-Primed Reactions 192

Figure 6.4. Accuracy vs. Base Position of All Reverse-Primed Reactions.....193

Figure 6.5. Average Error per 50-Base Window for Forward-Primed Reactions.....195

Figure 6.6. Average Error per 50-Base Window for Reverse-Primed Reactions 196

Figure 6.7. Deletion Errors per 50-Base Window for All Sequencing Reactions 197

Figure 6.8. Insertion Errors per 50-Base Window for All Sequencing Reactions 198

Figure 6.9. Miscall Errors per 50-Base Window for All Sequencing Reactions 199

Figure 6.10. No-Call Errors per 50-Base Window for All Sequencing Reactions 200

Figure 6.11. Errors of All Forward-Primed Reactions per 50-Base Window.....204

Figure 6.12. Errors of All Reverse-Primed Reactions per 50-Base Window 205

Figure 6.13. Color-Separation of Multiplets.....206

Figure 7.1. Sample Artifacts from PCR templates.....219

Figure 7.2. The Lump Artifacts in Two Different Sample Preparations.....221

Figure 7.3. The Stretc Artifact in Two Different Sample Preparations 222

Figure 7.4. Melting the Stretch Artifact with the Peltier Heater.....224

LIST OF TABLES

Table 2.1. Fluorescent Dyes and Bandpass Filters....39

Table 2.2. M13mp18 Cycle Sequencing ss Template DNA....48

Table 2.3. Wild-Type M13 Cycle Sequencing ds Template DNA....51

Table 2.4. APD BIAS at -20°C....66

Table 3.1. Refill Times of Various Fluids....76

Table 3.2. M13mp18....97

Table 3.3. Sample A....98

Table 3.4. Sample B....98

Table 3.5. Kepco Power Supply Settings for Peltier Heater....105

Table 3.6. Fluorescent Dyes and New Bandpass Filters....106

Table 3.7. Signal Change in Bases 74-77 Between Run #1 and #2....107

Table 3.8. APD Bias at -20.0°C....111

Table 4.1. Specifications of EG&G SiAPD at 22°C....136

Table 4.2. Calculated Specifications of SiAPD at -20°C....136

Table 4.3. Expected APD Detector Noise at -20°C....137

Table 4.4. Signal-to-Noise Ratios as Function of Fiber Optic....138

Table 4.5. LOD Values of Two Fiber Optics....138

Table 5.1. Size of Subcloned HSP60 PCR Fragments....164

Table 5.2. Dye-Labeled Primer Cycle-Sequencing Reaction....165

Table 5.3. GenBank Accession Numbers....167

Table 5.4. Related Organisms from BLAST....168

Table 5.5. Staphylococcus HSP60 Gene Genetic Distances and Sequence Similarities 169

Table 5.6. HSP60 Gene Genetic Distances and Sequence Similarities 170

Table 6.1. Error Distribution of Reactions per 50-Base Windows 202

Table 6.2. Common Deletion Error Sequences.....203

Table 6.3. Frequency of AAG and TTG Sequences in *Staphylococcus* Sequences.....203

Table 6.4. Frequency of Inserted Base.....208

Table 6.5. Most Common Miscall Errors 208

Table 6.6. Read Length at >99% Accuracy 210

Table 6.7. Read Length at >97% Accuracy. 210

Table 6.8. Read Length at >95% Accuracy 210

Table 7.1. Wild-Type M13 Cycle Sequencing Sample.....215

Table 7.2. Dye-Labeled Primer Cycle-Sequencing Reaction.....216

ABBREVIATIONS

A	amperes
APD	avalanche photodiode
CCD	charge-coupled device
CNS	coagulase-negative *Staphylococcus*
CZE	capillary zone electrophoresis
DMA	N,N-dimethylacrylamide
DNA	deoxyribonucleic acid
ds	double-stranded
EDTA	ethylenediaminetetraacetic acid
HV	high voltage
I	current
ID	inner diameter
GRIN	gradient index
HSP60	Heat Shock Protein-60 kDa
LED	light emitting diode
LOD	limit of detection
NA	numerical aperture
MCS	multiple cloning site
OD	outer diameter
PCR	polymerase chain reaction
PE/ABI	Perkin-Elmer Applied Biosystems
PID	Proportional/Integral/Derivative
PMT	photomultiplier tube

rRNA	ribosomal RNA
RTD	resistive temperature detector
SDS	sodium dodecyl sulfate
S/N	signal-to-noise
SPCM	single photon counting module
ss	single-stranded
TAE	tris-acetate-EDTA
TBE	tris-borate-EDTA
tris	tris(hydroxymethyl)aminomethane
TTE	tris-TAPS-EDTA
U	Units
V	Voltage
wt	wild-type
%T	total percent acrylamide in g/mL

$$= \frac{\text{mass}_{\text{acrylamide}} + \text{mass}_{\text{N'N'-methlylenebisacrylamide}}}{\text{volume}_{\text{water}}} \times 100$$

%C	percent cross-linked

$$= \frac{\text{mass}_{\text{N'N'-methlylenebisacrylamide}}}{\text{mass}_{\text{acrylamide}} + \text{mass}_{\text{N'N'-methlylenebisacrylamide}}} \times 100$$

Chapter 1: Introduction

1.1. INTRODUCTION

The Human Genome Project is an international initiative to sequence the entire human genome by the year 2005. The human genome is composed of 3 billion base pairs (10^9 base pairs = gigabase pairs = Gbp), with an estimated 50000 to 100000 genes [1]. The magnitude of this sequencing task is demonstrated by small prototypic projects. The *Escherichia coli* (*E. coli*) genome contains 4.6 million base pairs (10^6 base pairs = megabase pairs = Mbp). The determination of the *E. coli* genome required 7 years to complete [2]; 2.8 Mbp was sequenced in the last year of the project with the use of five PE Applied Biosystems slab gel sequencers [3]. Sequencing the 12 Mbp genome of the eukaryotic organism *Saccharomyces cerevisiae* required the collaboration of more than 600 scientists from over 100 laboratories [4].

From the inception of the Human Genome Program in 1990, there has been a need for at least a 10-fold increase in the current gene mapping, sequencing, and data analysis capabilities [1, 5, 6]. Large-scale DNA sequencing would require the development of automated, faster, and high-throughput technology to reduce time-consuming and labor-intensive steps.

Technological tools have improved since the beginning of the Genome Project. It is estimated that 1.2 Gbp of DNA sequence are in the public databases [7], including the complete genomes of *Haemophilus influenzae* (1.8 Mbp) [8], *Mycoplasma genitalium* (0.6 Mbp) [9], *Methanococcus jannaschii* (1.7 Mbp) [10], *Synechocystis* sp. PCC6803 (3.6 Mbp) [11], *Mycoplasma pneumoniae* (0.8 Mbp) [12], *Saccharomyces cerevisiae* (12 Mb) [4] and *Escherichia coli* (4.6 Mbp) [2]. However, only about 2% of the human genome has been sequenced to date [13]. In order to sequence the human genome by the 2005 year deadline, about 500 Mbp of sequence must be generated per year. The prokaryotic genome projects [8, 9] demonstrated that several Mbp of high accuracy sequence can be generated per month [14]. Clearly, there is a need to develop instrumentation that will allow faster DNA sequencing.

Capillary DNA sequencers have been developed to improve the speed and to increase the sample throughput of DNA sequencing [15-20]. Compared to current slab gel technology, capillary instruments are more sensitive and faster. For example, PE

Applied Biosystems' ABI Prism 310 capillary sequencer analyses 400 bases in 1 h and the ABI Prism 377 slab gel sequencer analyses 450 bases in 2.25 h [21, 22]. Capillary sample injection and gel refilling can be automated whereas slab gels require manual sample loading and gel replacement. Another advantage of capillary sequencers is their ability to analyze many samples simultaneously.

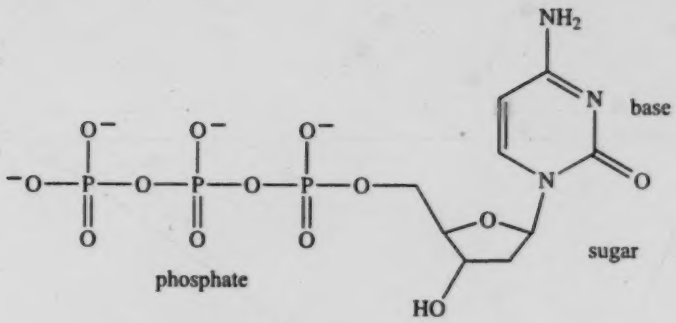
This thesis involves the development of a high-throughput capillary DNA sequencer and its use as a diagnostic tool to identify bacteria. A 16-capillary instrument was built based on a linear array of capillaries in a sheath-flow cuvette with laser-induced fluorescence detection. The instrument was evaluated by sequencing a portion of the Heat Shock Protein-60 kDa gene of several *Staphylococcus* species.

This introduction is broken into four sections. DNA sequencing is described in Section 1.2. In Section 1.3, multiple-capillary instrumentation is reviewed. In Section 1.4, the current methods to identify *Staphylococcus* species are presented. Finally, the main body of the thesis is outlined in Section 1.5.

1.2. DNA SEQUENCING

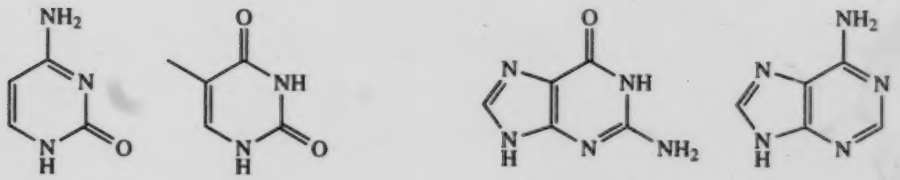
DNA sequencing is the determination of the order of the nucleotides: deoxyadenylate, deoxycytidylate, deoxyguanylate, and thymidylate. A nucleotide is composed of three units: the bases (adenine (A), cytosine (C), guanine (G), or thymine (T)), a deoxyribose sugar, and a phosphate group (Figure 1.1). Phosphodiester bonds join the 3' hydroxyl (-OH) of one sugar to the 5' hydroxyl of the next sugar forming a strand of DNA (Figure 1.2). Watson-Crick base-pairs are formed between the two DNA strands; A hydrogen bonds to T and C hydrogen bonds to G.

Sanger sequencing is the most frequently used technique [23]. DNA sequencing requires a DNA polymerase enzyme, a template, a primer, nucleotides (dNTPs, N = A, C, G, and T), and dideoxynucleotides (ddNTPs). The template is the DNA strand with the sequence of interest. A primer is a short piece of DNA that binds to its complementary sequence in the template. The DNA polymerase extends the 3' hydroxyl of the primer, adding nucleotides, which are Watson-Crick base-paired to the template sequence. The



deoxycytidine triphosphate (dCTP)

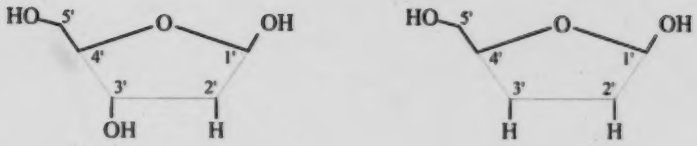
a) A nucleotide



cytosine thymine guanine adenine

Pyrimidines Purines

b) Bases



2'-deoxyribose 2',3'-dideoxyribose

c) Sugars

Figure 1.1. The Structure of DNA

A nucleotide is composed of a base and phosphate group connected to a sugar. The four bases of DNA are cytosine (C), adenine (A), guanine (G), and thymine (T).

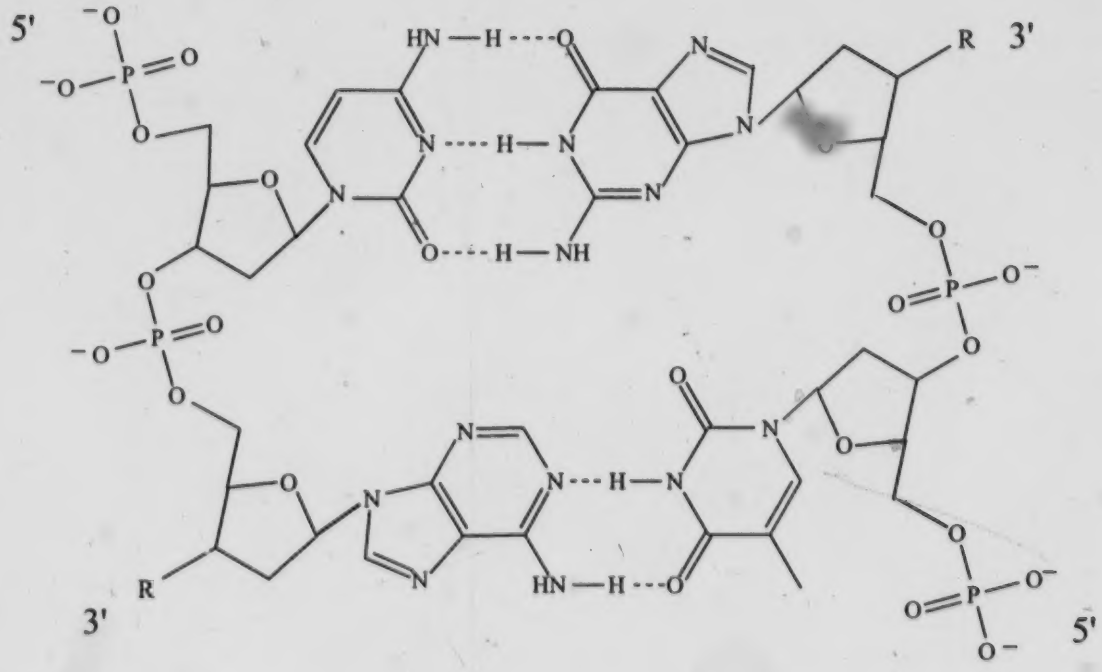


Figure 1.2. Base-Pairing of DNA

The four bases of DNA are: cytosine (C), adenine (A), guanine (G), and thymine (T). Phosphodiester bonds join the 3' position of one sugar to the 5' position of the next sugar. The DNA strand is synthesized from the 5' to 3' direction. Two strands of DNA run antiparallel and are joined by hydrogen bonds to form Watson-Crick base pairs; G pairs with C and A pairs with T.

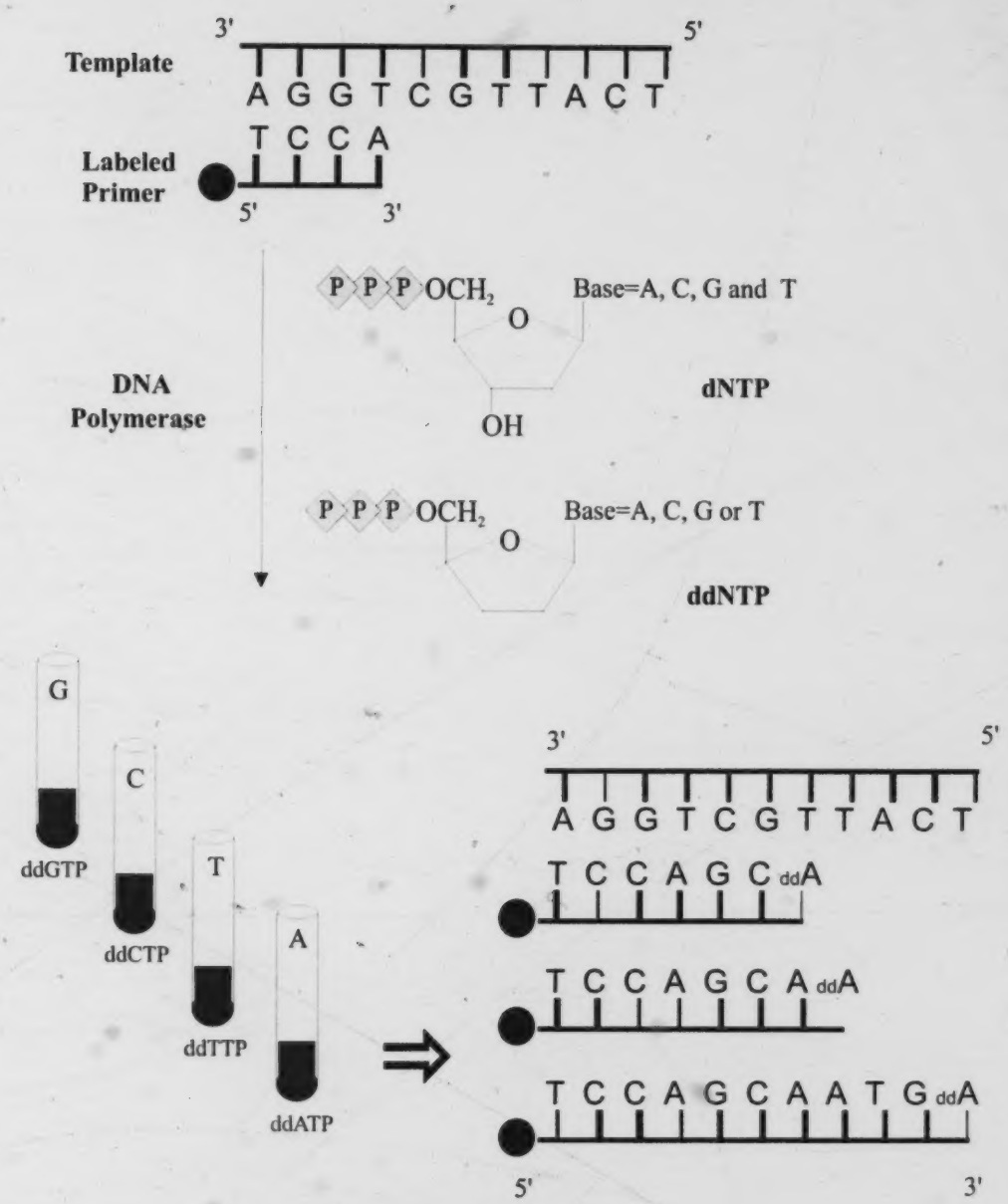


Figure 1.3. Sanger DNA Sequencing

A labeled primer hybridizes to a complementary region of a template according to Watson-Crick base-pairing: A binds to T and C binds to G. The enzyme, DNA polymerase, extends to 3' end of the primer with nucleotides, using the template as a map

DNA chain cannot be extended after a dideoxynucleotide is incorporated because it lacks a 3' hydroxyl group.

The DNA fragments are visualized by labeling the primer, nucleotides or dideoxynucleotides with a fluorescent tag. If the primer is labeled, the sequencing reactions are prepared in four vials (Figure 1.3). The DNA polymerase, template, labeled primer, and dNTPs are added to all the reaction vials. The fluorescent tag on the primer is different in each vial and only one of the ddNTPs is added to a vial. For example, in the A vial, only ddATP is added to the sequencing reagents and A-terminated fragments are synthesized. After the reactions, the four vials are then combined and separated by size using gel electrophoresis. Alternatively, the same fluorescently labeled primer can be used in all four reactions and the samples separately run on four lanes in the gel.

A description of Polymerase Chain Reaction (PCR) and DNA cycle sequencing can be found in undergraduate biochemistry textbooks [24, 25].

1.3. MULTIPLE-CAPILLARY INSTRUMENTS

Zagursky and McCormick at du Pont reported the first DNA sequencing with multiple capillaries [26]. They placed twelve 530 μm ID capillaries in a modified GENESIS 2000 slab gel machine. Fluorescence was detected on-column. An argon ion laser beam is scanned along the capillaries and the fluorescence was detected at two wavelengths by two photomultiplier tubes (PMTs). They used four different fluorescent labels for each DNA base. The ratio of fluorescence in the two spectral channels identified the DNA sequence. They sequenced 500 bases per capillary in 9.5 hours with an accuracy of 96%.

Mathies' group at University of Berkeley described a capillary array electrophoresis instrument [27]. The instrument was a confocal, on-column, one-color, capillary array scanner. The 25-capillary array moved underneath the optical system at a rate of 2 cm/s and each capillary was sequentially analyzed. The instrument had an epi-illumination geometry; the laser excitation light was focused on the capillary by a microscope objective and fluorescence emission was collected by the same objective. A beamsplitter reflected the laser line and transmitted the fluorescence directly to a PMT.

The authors did not find a significant decrease in sensitivity of their instrument between scanning-mode and stationary mode; they reported a limit of detection (LOD) of 2×10^{-12} M and 1×10^{-12} M of fluorescein (signal-to-noise ratio, S/N = 3) for scanning and stationary modes respectively.

An alteration to the 25-capillary array design was later reported involving 2-color detection with two PMTs and a binary-coding scheme to identify the DNA bases [28, 29]. FAM and JOE are fluorescent dyes attached to the sequencing primers and the DNA bases were labeled as follows: A-FAM and JOE; C-no dye; G-JOE; and T-FAM. The absence of a dye for C lead to high errors in the sequencing run; an accuracy of 94.6% was obtained in the first 280 bases. A more recent paper from the same group reported a four-color array instrument with four PMTs [30]. The 25-capillary instrument was effectively only 12-capillary because an alignment capillary containing fluorescein dye, presumably for lane tracking, was placed between each capillary. The authors reported an improved sequencing accuracy of 99.7% (after-editing) for sequences approximately 470 bases after the primer. Molecular Dynamics is currently developing this instrument design [31].

The biggest disadvantage of the Mathies' multiple capillary instrument design is the sequential data acquisition, which limits the number of capillaries that can be scanned with high sensitivity. As the number of capillaries increases, the duty cycle decreases. A low duty cycle results in a decrease in optical collection efficiency and a deterioration in the limit of detection of the instrument. In sequential scanning, a low duty cycle results from the short time each capillary is probed. The duty cycle is 4% for a 25-capillary instrument and 1% for a 100-capillary instrument.

Yeung's group at Iowa State University investigated several different on-column detection geometries for multiple-capillary electrophoresis. One design used axial-beam excitation, where the laser illumination was transmitted by a fiber optic inserted into the end of the capillary [32]. The fluorescence from all the capillaries was simultaneously collected at right-angles by a charge-coupled device (CCD) camera positioned over the capillary array. The disadvantage of this instrument design was the placement of the fiber optic in the capillary. The fiber optic disrupted electrophoresis and made gel refilling

impractical between runs. The cross-talk between neighboring capillaries was reduced by separating each capillary with a capillary coated in black.

Another instrument from Yeung's group used an elliptically shaped beam, at a 45° angle to the capillary array, to simultaneously excite 100 capillaries [33]. A CCD camera was positioned perpendicularly to the array. This design was more sensitive than the axial-excitation design with LODs of 4×10^{-10} to 8×10^{-10} M fluorescein (S/N = 3). The cross-talk was reduced to 10% by placing capillaries in aluminum grooves.

However, the most recent design from Yeung's group has the best LOD of their instrument designs. The sequencer utilized side-on laser illumination of 9 capillaries immersed in an refractive-index matching solution to reduce scatter and cross-talk [34]. Yeung reported an LOD at the low pM (1×10^{-12} M) level, which was 2 orders of magnitude improvement over their previous designs.

Quesada at Brookhaven National Laboratories built an 8-capillary instrument that used a fiber optic to deliver the laser excitation beam and a second perpendicular fiber optic to collect fluorescence emission [35]. An LOD of 1.5×10^{-11} M fluorescein (S/N = 2) was reported for open-tube capillaries. The advantages of this design are simultaneous excitation and signal collection. However, capillary replacement and alignment appears to be difficult; the fiber optics were permanently glued to the capillary surface. The design is currently limited to the number of fiber optics connected to commercially available spectrographs.

Kambara's group at Hitachi built a 20-capillary array instrument using post-column sheath-flow detection [36, 37]. The laser excitation beam skimmed under the capillary tips. The fluorescence was collected from all capillaries simultaneously at 90° angle by a cylindrical lens and focused onto a CCD camera. Sheath-buffer flowing down the sides of the capillary focused the DNA stream. The DNA traveled 1 mm and flowed into a lower open-tube capillary. In order to align the separation capillaries (100 µm ID, 200 µm OD) with the lower capillaries (200 µm ID, 350 µm OD), the capillaries were spaced with a periodicity of 350 µm. Kambara reported an LOD of 1×10^{-13} M of Texas Red in 1-color mode and 2×10^{-12} M (S/N = 2) in 4-color mode with the PE Applied Biosystems' dyes FAM, JOE, TAMRA and ROX.

Kambara's group also built an on-column 24-capillary instrument [38]. In this design, the laser simultaneously excited all the capillaries in a linear array. Although the laser light was scattered and refracted by each capillary, it was refocused by the lens-effect attributed to capillary curvature. They reported an average transmittance of 91% per capillary, which will limit the total number of capillaries that can be used. Their signal-to-noise ratios dropped from 1000 at capillary 1 to 700 at capillary 10. They reported an LOD of ~1×10^{-12} M Texas Red from flowing 10^{-9} M solution through the first 10 capillaries.

Dovichi's group at the University of Alberta built a 5-capillary instrument using post-column sheath-flow detection [39]. The capillaries, 50 μm ID, 150 μm OD, were fit into a rectangular quartz cuvette. The sides of the cuvette tapered to hold the capillaries in a linear array of 150 μm pitch. The laser beams skimmed under the capillary tips and fluorescence was collected by a microscope objective placed at right-angles to the excitation. The fluorescence signal was passed through a chopper wheel and imaged onto avalanche photodiodes. Zhang reported an LOD of 2×10^{-13} M fluorescein (S/N = 3), which was an order of magnitude better than the lowest reported LOD for a multiple-capillary instrument.

The key to the sensitivity of Dovichi's capillary instrumentation lies in the post-column sheath-flow cuvette [40]. In the on-column capillary instruments, the sample is excited while migrating through the capillary. This design suffers from high backgrounds from laser scatter. The scatter arises from the large change in refractive index at the capillary-sample interface. This laser scatter is significantly reduced by the sheath-flow cuvette design. Sheath-fluid flows from the top of the cuvette, and focuses the flow stream of fluorescently labeled sample as it exits the capillary. The laser is positioned under the capillary tips to excite the sample. The sample flow streams have the identical refractive index with surrounding sheath buffer; there is no light scatter at the sample stream-sheath buffer interface. Laser scatter is also reduced as the laser enters the rectangular cuvette, which has optically flat walls.

1.4. *STAPHYLOCOCCUS* IDENTIFICATION

There are about 30 identified species in the *Staphylococcus* genus, half are indigenous to humans, and within each species there are many strains [41]. The species of the *Staphylococcus* genus are distinguished from other micrococci based on phenotypic tests. For example, *Staphylococci* are Gram-positive, cluster-forming, and able to produce acid anaerobically from glucose [42]. *Staphylococcus* species are characterized by their production of the enzyme coagulase, which promotes fibrin clotting. If cells are coagulase-positive, they are usually identified as the species *S. aureus,* although coagulase-positive species also include *S. intermedius*, *S. delphini* and *S. schleiferi* subsp. *coagulans,* which are major human pathogens [43]. Coagulase-negative bacteria are simply reported as coagulase-negative *Staphylococcus* species (CNS). CNS isolates are composed of many different species and strains.

The majority of the bacteria found on human skin are CNS and CNS are one of the most frequently isolated bacteria in hospital microbiology laboratories. The pathogenic strains are assumed to cause the majority of nosocomial, hospital-acquired, infections [41]. Nosocomial infections arise from the increased use of medical devices such as prosthetics and intravascular catheters that allow CNS to gain entry into their host. It is important for clinical laboratories to differentiate the clinically significant and pathogenic *Staphylococcus* species from normal contaminating *Staphylococcus* species quickly and accurately for proper patient treatment.

Conventional methods to identify the *Staphylococcus* species and strains rely on growing the bacteria in culture. A major disadvantage of culturing is that bacteria can take days to grow. The specimen is collected from the patient (e.g. blood or sputum) and a colony or group of colonies is selected, grown and examined phenotypically, for an visible characteristic, or genotypically, based on the DNA sequence.

Phenotypic tests include colony morphology (color, size, luster) and biotyping (determining antibiotic resistance and plasmids in bacteria) [42, 44]. Positive or negative enzyme activities (e.g. alkaline phosphatase, urease, and β-galactosidase) and the ability to aerobically metabolize certain sugars into acids can also identify *Staphylococcus* species and strains. Combining these tests successfully identifies *Staphylococcus* strains.

The tests, however, are quite laborious and time-consuming, taking up to 5 days for identification [45].

There are several commercially available kits that use these biochemical reactions to identify *Staphylococcus* species and strains in a few hours to one day. These kits are able to identify most species with accuracies ranging from 70 to 100% [41, 46-50]. The kits are expensive, have poorer accuracies (<60%) for more recently discovered and novel *Staphylococcus* species, and even misidentify common species [46-50]. Presumably the accuracy of these kits will improve as the database of species grows.

Other phenotypic approaches are being developed to identify *Staphylococci* more rapidly, reproducibly, affordably, and with higher sensitivity and specificity. Cellular fatty acid analysis [51, 52] reagents and instruments are relatively inexpensive and correlate well with the results from conventional techniques. Pyrolysis-mass spectrometry of the bacteria is inexpensive, after instrument purchase, rapid and reported to be reproducible [53]. Metabolite enzyme electrophoresis profiles [54, 55] are reproducible and simple but expensive. Whole-cell peptide analysis by SDS polyacrylamide gel electrophoresis and visualization with stains or immunoblotting [41] is sensitive and fast. However without the proper antigen-antibody recognition, new species cannot be identified.

Genotypic nucleic acid sequence analysis examines either plasmid DNA or chromosomal DNA profiles. Plasmid profiling through restriction fragment length polymorphism (RFLP) fingerprinting is inexpensive and simple [56]. Many plasmids, however, are highly conserved in strains and they have identical fingerprint patterns between strains. Plasmids are also lost or added, reducing the ability to identify strains. Chromosomal RFLP analysis by pulsed-field gel electrophoresis is more discriminating than plasmid fingerprinting in identifying strains [55, 57].

More recently, genotypic identification methods focus on gene sequences that are either unique to a bacterial species or ubiquitous to all bacterial species but contain regions of variability within the genes. Polymerase chain reaction (PCR) methods are used to identify genes unique to antibiotic resistant *Staphylococcus* strains. The *mecA*

gene in *S. aureus* is a genetic determinant for resistance to antibiotic methicillin. PCR with primers specific for *mecA* gene identified the methicillin resistant strain [58-60].

Ribotyping involves restriction enzyme cutting of the chromosomal DNA and then probing with radiolabeled 16S ribosomal RNA (rRNA) gene from *Bacillus subtilis* [61, 62] or *E.coli* [63, 64]. The fingerprint pattern identifies species and even differentiates strains in some cases.

DNA sequencing of the rRNA gene directly is a more precise method to identify bacteria. Currently, rRNA gene sequences are the most common source of information to identify, classify and trace evolutionary pathways. The Ribosomal Database Project (http://www.cme.msu.edu/RDP) [65] has over 11000 organisms listed.

Chow proposed a method similar to rRNA gene sequencing. The heat shock protein 60-kDa (HSP60), also known as GroEL, is a highly conserved, single copy gene [66]. PCR amplification of a portion of this gene with a single set of primers resulted in a 600 bp PCR product. The same primer set amplified a 600 bp PCR product from all tested *Staphylococcus* species. The PCR products from different *Staphylococcus species* did not cross hybridize under stringent conditions indicating regions of variable sequence within the highly conserved gene. As a result, sequencing their HSP60 PCR product can identify different bacterial species.

In addition to species identification, genus identification is possible. The HSP60 PCR primers amplified a 600 bp product from other Gram-positive microbes: *Bacillus subtilis, Streptococcus agalactiae,* and *Mycobacterium marinum.* Gram-negative bacteria were also amplified: *Salmonella typhi, Escherichia coli, Haemophilus influenzae*, and *Helicobacter pylori* [66].

PCR is so sensitive that it can amplify as little as one molecule of DNA. The advantage of the PCR technique is that it may circumvent the requirement for bacterial growth in cultures. Instead, PCR may identify bacteria *in situ*, at the focus of the infection, where bacteria are not normally present (e.g. cerebrospinal fluid). PCR would then reduce the entire analysis time to hours instead of the days required to grow and isolate bacteria from cultures. Combined with capillary DNA sequencing, the identification of bacteria based on DNA sequence could be performed in hours.

Contamination is a problem in cultures and PCR if the specimen is not collected aseptically. To date, PCR has only been performed on bacteria isolated from cultures; it remains to be seen if PCR can be performed directly from patient specimens. It is uncertain if the HSP60 sequences will indicate the pathogenicity of the bacteria or discriminate strains within species of *Staphylococcus*.

1.5. THIS THESIS

Spurred by the Human Genome Project, the 16-capillary instrument was developed as a high-throughput DNA sequencer. The instrument was used as a diagnostic tool to hasten genotypic identification of bacteria. This thesis presents the implementation of high-throughput DNA sequencing from instrument design, construction and evaluation to sample preparation and data analysis.

The construction of the first generation 16-capillary DNA sequencer is described in Chapter 2. A micromachined sheath-flow cuvette was fabricated for laser-induced fluorescence detection. Glass structures inside the cuvette ensured the reproducible alignment of the sample streams from the 16 capillaries to the detection optics. Four-color DNA samples were spectrally discriminated through a spinning filter wheel. Avalanche photodiode detectors detected the fluorescent signal with high-gain external amplifiers. The instrument was evaluated by running samples of known DNA sequence.

Chapter 3 details the second generation 16-capillary sequencer. The avalanche photodiode detectors were optimized to improve the signal-to-noise ratio of the instrument. The separation apparatus of the instrument was modified to facilitate its use. A new sample injector injects 16 samples simultaneously. A pressurized capillary gel refiller was adapted to the setup for parallel refilling of sixteen capillaries.

Chapter 4 outlines the procedure to optimize the signal-to-noise ratio of the avalanche photodiodes detectors. By decreasing the noise and increasing the signal of the detectors, the APD detectors were shown to be signal-shot-noise-limited, a fundamental noise limiting situation. A limit of detection value was reported for the 16-capillary instrument.

In Chapter 5, the second generation 16-capillary sequencer performance was evaluated by sequencing DNA samples of unknown sequence. A set of reference DNA sequences were determined for several *Staphylococcus* species. A phylogenetic tree of the *Staphylococcus* species was constructed using the partial sequence of the Heat Shock Protein 60 kDa gene.

In Chapter 6, the DNA base-calling software BASS was evaluated for sequencing accuracy using the *Staphylococci* runs. The errors were broken into four types: deletion, insertion, miscall and no-call. The average read lengths were determined based on sequencing accuracy.

Chapter 7 details some sequencing artifacts that ruined sequencing accuracy and artifacts that were catastrophic to certain sequencing runs. Sample preparation protocols were altered to clean templates and a Peltier capillary heater was employed to remove sample artifacts from sequencing runs.

Chapter 8 presents conclusions about the thesis and proposes future work.

1.6. REFERENCES

(1) Hunkapiller, T.; Kaiser, R. J.; Koop, B. F.; Hood, L. *Science* **1991,** *254*, 59-67.

(2) Blattner, F. R.; Plunkett, G. r.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D.; Mau, B.; Shao, Y. *Science* **1997,** *277*, 1453-74.

(3) O'Brien, C. *Nature* **1997,** *385*, 472.

(4) Mewes, H. W.; Albermann, K.; Bahr, M.; Frishman, D.; Gleissner, A.; Hani, J.; Heumann, K.; Kleine, K.; Maierl, A.; Oliver, S. G.; Pfeiffer, F.; Zollner, A. *Nature* **1997,** *387*, 7-65.

(5) Trainor, G. L. *Analytical Chemistry* **1990,** *62*, 418-426.

(6) Olsen, M. V. *Analytical Chemistry* **1991,** *63*, 416-20A.

(7) http://www.ncbi.nlm.nih.gov, National Center for Biotechnology Information, 1998.

(8) Fleischmann, R. D.; Adams, M. D.; White, O.; Clayton, R. A.; Kirkness, E. F.; Kerlavage, A. R.; Bult, C. J.; Tomb, J. F.; Dougherty, B. A.; Merrick, J. M. *Science* **1995,** *269*, 496-512.

(9) Fraser, C. M.; Gocayne, J. D.; White, O.; Adams, M. D.; Clayton, R. A.; Fleischmann, R. D.; Bult, C. J.; Kerlavage, A. R.; Sutton, G.; Kelley, J. M. *Science* **1995,** *270*, 397-403.

(10) Bult, C. J.; White, O.; Olsen, G. J.; Zhou, L.; Fleischmann, R. D.; Sutton, G. G.; Blake, J. A.; FitzGerald, L. M.; Clayton, R. A.; Gocayne, J. D.; Kerlavage, A. R.; Dougherty, B. A.; Tomb, J. F.; Adams, M. D.; Reich, C. I.; Overbeek, R.; Kirkness, E. F.; Weinstock, K. G.; Merrick, J. M.; Glodek, A.; Scot, t. J. L.; Geoghagen, N. S. M.; Venter, J. C. *Science* **1996,** *273*, 1058-73.

(11) Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S. *DNA Research* **1996,** *3*, 109-36.

(12) Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B. C.; Herrmann, R. *Nucleic Acids Research* **1996,** *24*, 4420-49.

(13) Beardsley, T. In *Scientific American*, 1998; Vol. February, pp 28-29.

(14) Venter, J. C.; Smith, H. O.; Hood, L. *Nature* **1996,** *381*, 364-6.

(15) Swerdlow, H.; Gesteland, R. *Nucleic Acids Research* **1990,** *18*, 1415-1419.

(16) Swerdlow, H.; Wu, S.; Harke, H.; Dovichi, N. J. *Journal of Chromatography* **1990,** *516*, 61-67.

(17) Luckey, J. A.; Drossman, H.; Kostichka, A. J.; Mead, D. A.; D'Cunha, J.; Norris, T. B.; Smith *Nucleic Acids Research* **1990,** *18*, 4417-4421.

(18) Heiger, D. [redacted] Cohen, A. S.; Karger, B. L. *Journal of Chromatography* **1990,** *516*, 49-60.

(19) Drossman, H.; Luckey, J. A.; Kostichika, A. J.; D'Cunha, J.; Smith, L. M. *Analytical Chemistry* **1990,** *62*, 900-903.

(20) Swerdlow, H.; Zhang, J. Z.; Chen, D. Y.; Harke, H. R.; Grey, R.; Wu, S.; Dovichi, N. J.; Fuller, C. *Analytical Chemistry* **1991,** *63*, 2835-2841.

(21) ABI PRISM 310 Genetic Analyzer, Perkin-Elmer/Applied Biosystems, **1997**.

(22) ABI PRISM 377 DNA Sequencer, Perkin-Elmer/Applied Biosystems, **1997**.

(23) Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. USA* **1977,** *74*, 5463-67.

(24) Brown, T. A. *Gene Cloning: An Introduction*; Chapman and Hall: New York, **1995**.

(25) Mathews, C. K.; van Holde, K. E. *Biochemistry*; Benjamin/Cummings Publishing Company, Inc.: Don Mills, Ontario, **1990**.

(26) Zagursky, R. J.; McCormick, R. M. *BioTechniques* **1990,** *9*, 74-79.

(27) Huang, X. C.; Quesada, M. A.; Mathies, R. A. *Analytical Chemistry* **1992,** *64*, 967-972.

(28) Huang, X. C.; Quesada, M. A.; Mathies, R. A. *Analytical Chemistry* **1992,** *64*, 2149-2154.

(29) Mathies, R. A.; Huang, X. C. *Nature* **1992,** *359*, 167-169.

(30) Kheterpal, I.; Scherer, J. R.; Clark, S. M.; Radhakrishnan, A.; Ju, J.; Ginther, C. L.; Sensabaugh, G. F.; Mathies, R. A. *Electrophoresis* **1996,** *17*, 1852-1859.

(31) Bashkin, J. S.; Bartosiewicz, M.; Roach, D.; Leong, J.; Barker, D.; Johnston, R. *Journal of Capillary Electrophoresis* **1996,** *3*, 61-68.

(32) Taylor, J. A.; Yeung, E. A. *Analytical Chemistry* **1993,** *65*, 956-960.

(33) Ueno, K.; Yeung, E. S. *Analytical Chemistry* **1994,** *66*, 1424-1431.

(34) Lu, X.; Yeung, E. S. *Applied Spectroscopy* **1995,** *49*, 605-609.

(35) Quesada, M. A.; Zhang, S. *Electrophoresis* **1996,** *17*, 1841-1851.

(36) Kambara, H.; Takahashi, S. *Nature* **1993,** *361*, 565-566.

(37) Takahashi, S.; Murakami, K.; Anazawa, T.; Kambara, H. *Analytical Chemistry* **1994,** *66*, 1021-1026.

(38) Anazawa, T.; Takahashi, S.; Kambara, H. *Analytical Chemistry* **1996,** *68*, 2699-2704.

(39) Zhang, J. Z. Z. Ph.D., University of Alberta, Edmonton, **1994**.

(40) Cheng, Y. F.; Wu, S.; Chen, D. Y.; Dovichi, N. J. *Analytical Chemistry* **1990,** *62*, 496-503.

(41) Kloos, W. E.; Bannerman, T. L. *Clinical Microbiology* **1994,** *7*, 117-140.

(42) Baird-Parker, A. C. In *The Staphylococci*; Cohen, J. O., Ed.; John Wiley and Sons, Inc.: Toronto, Ontario, **1972**, pp 1-20.

(43) Kloos, W. E.; Lambe, D. W. In *Manual of Clinical Microbiology*; Balows, A., Hausler, W. J., Herrmann, K. L., Isenberg, H. D., Shadomy, H. J., Eds.; American Society for Microbiology: Washington, D.C., **1991**, pp 222-237.

(44) Richmond, M. H. In *The Staphylococci*; Cohen, J. O., Ed.; John Wiley and Sons, Inc.: Toronto, Ontario, **1972**, pp 159-186.

(45) Kleeman, K. T.; Bannerman, T. L.; Kloos, W. E. *Journal of Clinical Microbiology* **1993,** *31*, 1318-1321.

(46) Bannerman, T. L.; Kleeman, K. T.; Kloos, W. E. *Journal of Clinical Microbiology* **1993,** *21*, 1322, 1325.

(47) Kellogg, J. A., Hanna, M.D., Nelsem, S.J., Sprenkle, L.S., Thomas, M.L., Young, K.S. *Clinical Microbiology and Infectious Disease* **1996,** *106*, 374-377.

(48) Janda, W. M.; Ristow, K.; Novak, D. *Journal of Clinical Microbiology* **1994,** *32*, 2056-2059.

(49) Renneberg, J.; Rieneck, K.; Gutschik, E. *Journal of Clinical Microbiology* **1995,** *33*, 1150-1153.

(50) Ieven, M.; Verhoeven, J.; S.R., P.; Goosens, H. *Journal of Clinical Microbiology* **1995,** *33*, 1060-1063.

(51) Welch, D. *Clinical Microbiology Reviews* **1991,** *4*, 422-438.

(52) Birnbaum, D.; Herwaldt, L.; Low, D. E.; Noble, M.; Pfaller, M.; Sherertz, R.; Chow, A. W. *Journal of Clinical Microbiology* **1994,** *32*, 2113-2119.

(53) Freeman, R., Goodfellow,M., Ward, A.C., Hudson, S.J., Gould. F.K., Lightfoot, N.F. *Journal of Medical Microbiology* **1991,** *34*, 245-248.

(54) Pfaller, M. A. *Hoechst Clinical Microbiology* **1992,** *4*, 1-7.

(55) Schlichting, C.; Branger, C.; Fournier, J.-M.; Witte, W.; Boutonnier, A.; Wolz, C.; Goullet, P.; Doring, G. *Journal of Clinical Microbiology* **1993,** *31*, 227-232.

(56) Pfaller, M. A.; Hollis, R. J. *Clinical Microbiology Newsletter* **1989,** *11*, 137-142.

(57) Linhardt, F.; Ziebuhr, W.; Meyer, P.; Witte, W.; Hacker, J. *FEMS Microbiology Letters 95* **1992**, 181-186.

(58) Murakami, K.; Minamide, W.; Wada, K.; Nakamura, E.; Teraoka, H.; Watanabe, S. *Journal of Clinical Microbiology* **1991,** *29*, 2240-2244.

(59) Geha, D. J., Uhl, J.R., Gustaferro, C.A., Persing, D.H. *Journal of Clinical Microbiology* **1994,** *32*, 1768-1772.

(60) Vannuffel, P.; Gigi, J.; Ezzendine, H.; Vandercam, B.; Delmee, M.; Wauters, G.; Gala, J.-L. *Journal of Clinical Microbiology* **1995,** *33*, 2864-2867.

(61) DeBuyser, M.-L.; Morvan, A.; Grimont, F.; El Sohl, N. *Journal of General Microbiololgy* **1989,** *135*, 989-999.

(62) Chesneau, O.; Aubert, S.; Morvan, A.; Guesdon, J. L.; El Solh, N. *Journal of Clinical Microbiology* **1992,** *30*, 2346-2352.

(63) Thomson-Carter, F. M.; Carter, P. E.; Pennington, T. H. *Journal of General Microbiology* **1989,** *135*, 1093-2097.

(64) Webster, J. A.; Bannerman, T. L.; Hubner, R. J.; Ballard, D. N.; Cole, E. M.; Bruce, J. L.; Fiedler, F.; Schubert, K.; Kloos, W. E. *International Journal of Systematic Bacteriology* **1994,** *44*, 454-460.

(65) Maidak, B. L.; Olsen, G. J.; Larsen, N.; Overbeek, R.; McCaughey, M. J.; Woese, C. R. *Nucleic Acids Research* **1997,** *25*, 109-11.

(66) Goh, S. H.; Potter, S.; Wood, J. O.; Hemmingsen, S. M.; Reynolds, R. P.; Chow, A. W. *Journal of Clinical Microbiology* **1996,** *34*, 818-823.

Chapter 2: The 16-Capillary Sequencer, Version 1.0

2.1. INTRODUCTION

I introduce here the first generation of a 16-capillary DNA sequencer based on a linear array of capillaries held in a micromachined sheath-flow cuvette. The description of the 16-capillary instrument is broken into two parts: the separation apparatus and the detector apparatus. A schematic of the separation apparatus is shown in Figure 2.1. It consisted of the micromachined sheath-flow cuvette, 16 gel-filled capillaries, a capillary heater, and a sample injector. An overall schematic of the detector apparatus is shown in Figure 2.2. It consisted of two lasers, the collection optics with Gradient Index (GRIN) lenses, the avalanche photodiodes (APDs), the chopper/filter wheels, and the data-acquisition. The instrument was evaluated by sequencing standard bacteriophage DNA and cloned DNA, both of known sequence.

2.1.1. Separation Apparatus

Glass Micromachining

The 16-capillary instrument used a micromachined glass cuvette. The basic steps of photolithography are illustrated in Figure 2.3. A planar substrate, such as silicon or glass, is coated with a thin metal layer. The metal layers are coated with a photosensitive polymer, called photoresist. A mask design that is analogous to a negative in photography is positioned over the substrate. The photoresist is exposed to UV light through the mask and areas of photoresist exposed to UV light are removed after developing. The exposed metal is removed and the exposed substrate areas are etched to the desired depth, while areas covered by the photoresist are not etched. In the final step, the remaining photoresist and metal layers are removed.

Glass is composed of silicon dioxide (SiO_2) and etched with hydrofluoric acid (HF). A proposed reaction is [1]:

$$SiO_2 + 6HF \rightarrow H_2SiF_6 + 2H_2O$$

Glass is amorphous, with no crystalline structure, and wet chemical etches with HF are isotropic; the substrate is etched in all directions. In contrast, anisotropic etches ("orientation dependent") are limited by the crystallographic plane angles in a single silicon crystal. The etched features have sharply defined angles in the substrate [2].

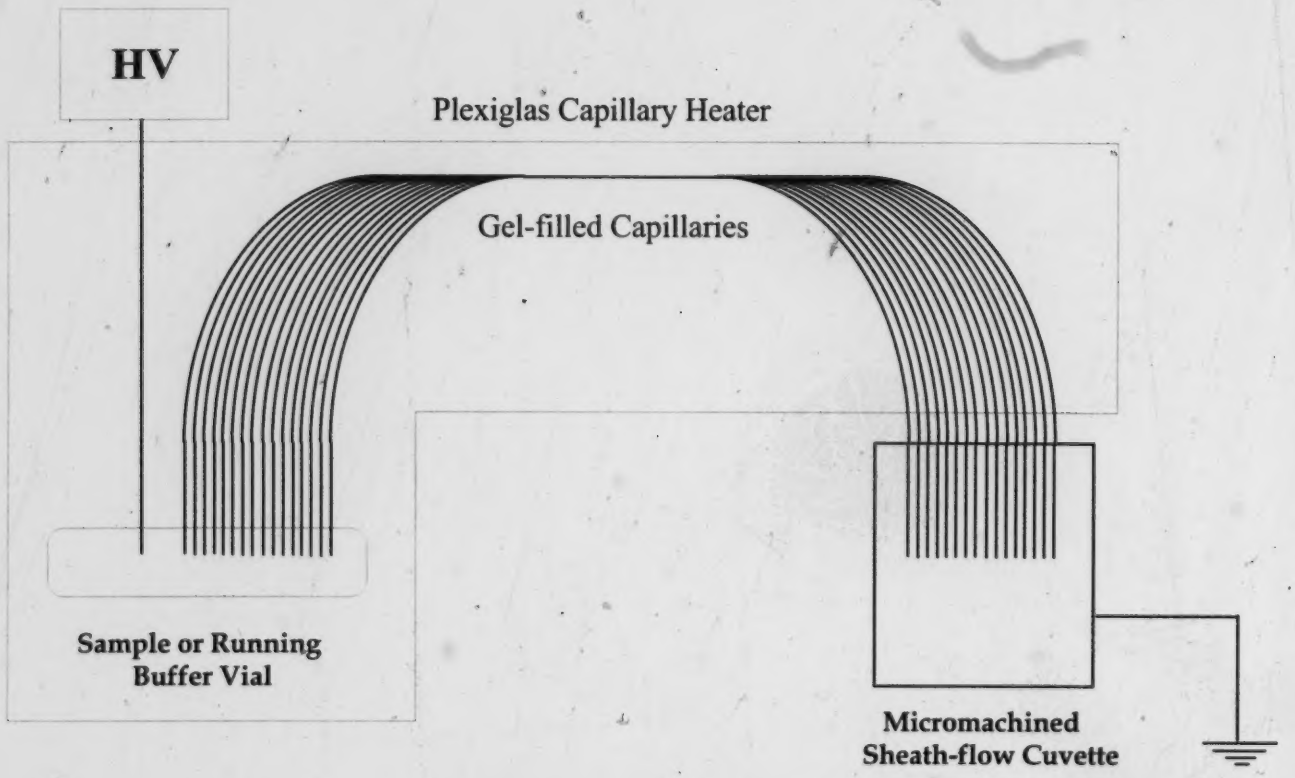


Figure 2.1. The Separation Apparatus.
Sixteen gel-filled capillaries are inserted in a linear array in the cuvette. The other ends of the capillaries are dipped in the DNA sample, which is loaded onto the capillaries by briefly applying a high voltage. The capillary ends are then dipped into a vial containing buffer and electrophoresis is continued. The negative voltage forces the negatively-charged DNA fragments to the cuvette end at ground potential. A Plexiglas heater with forced air thermostats the capillaries.

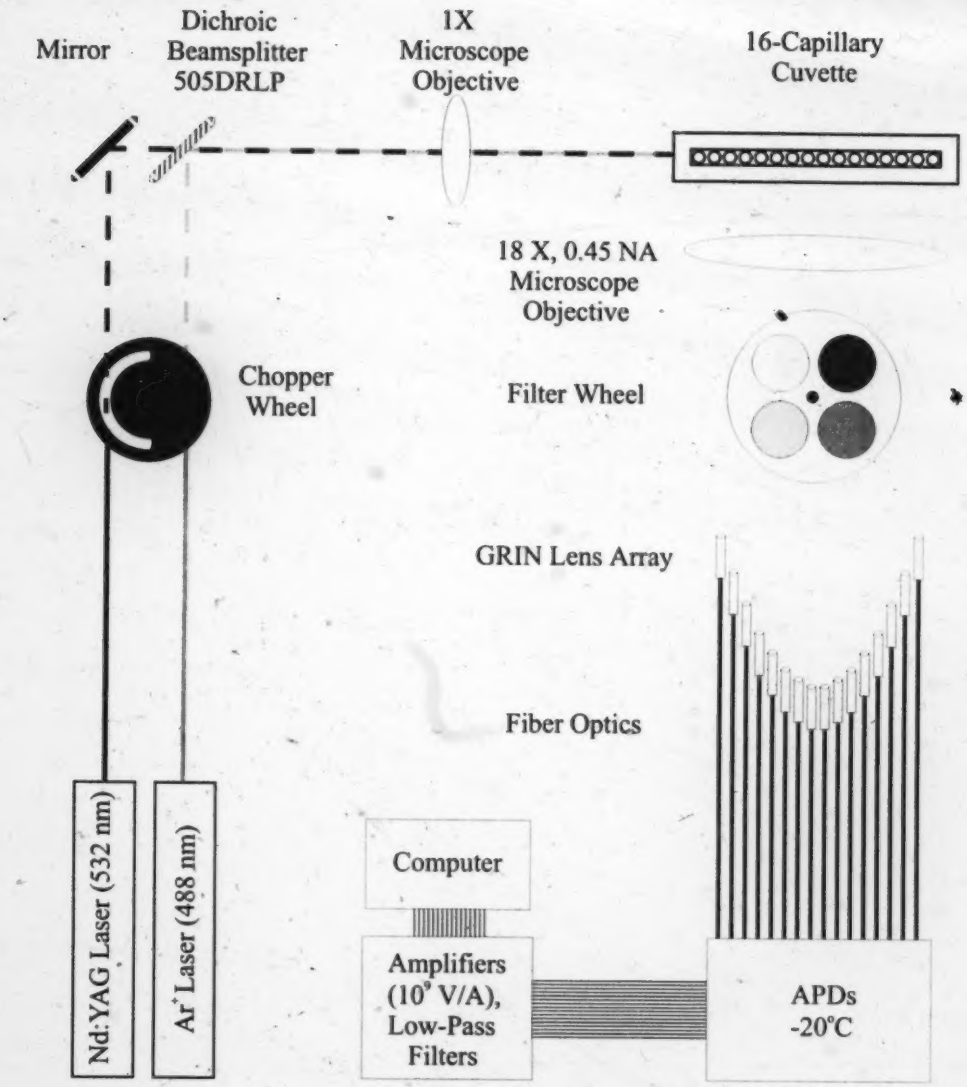


Figure 2.2. The Detector Apparatus.
The laser beams are chopped by the chopper wheel and recombined by the mirror and dichroic beamsplitter. The laser beams are focused onto the cuvette by a microscope objective. The laser beams excite the fluorescently-labeled DNA as it exits the capillary. The fluorescence is collected at right angles to the laser beams by a microscope objective, passed through a filter wheel and imaged onto the GRIN lens array. Fiber optic cables transmit the fluorescence to the APDs. The light is converted to current, amplified, converted to voltage, filtered, and read into the computer.

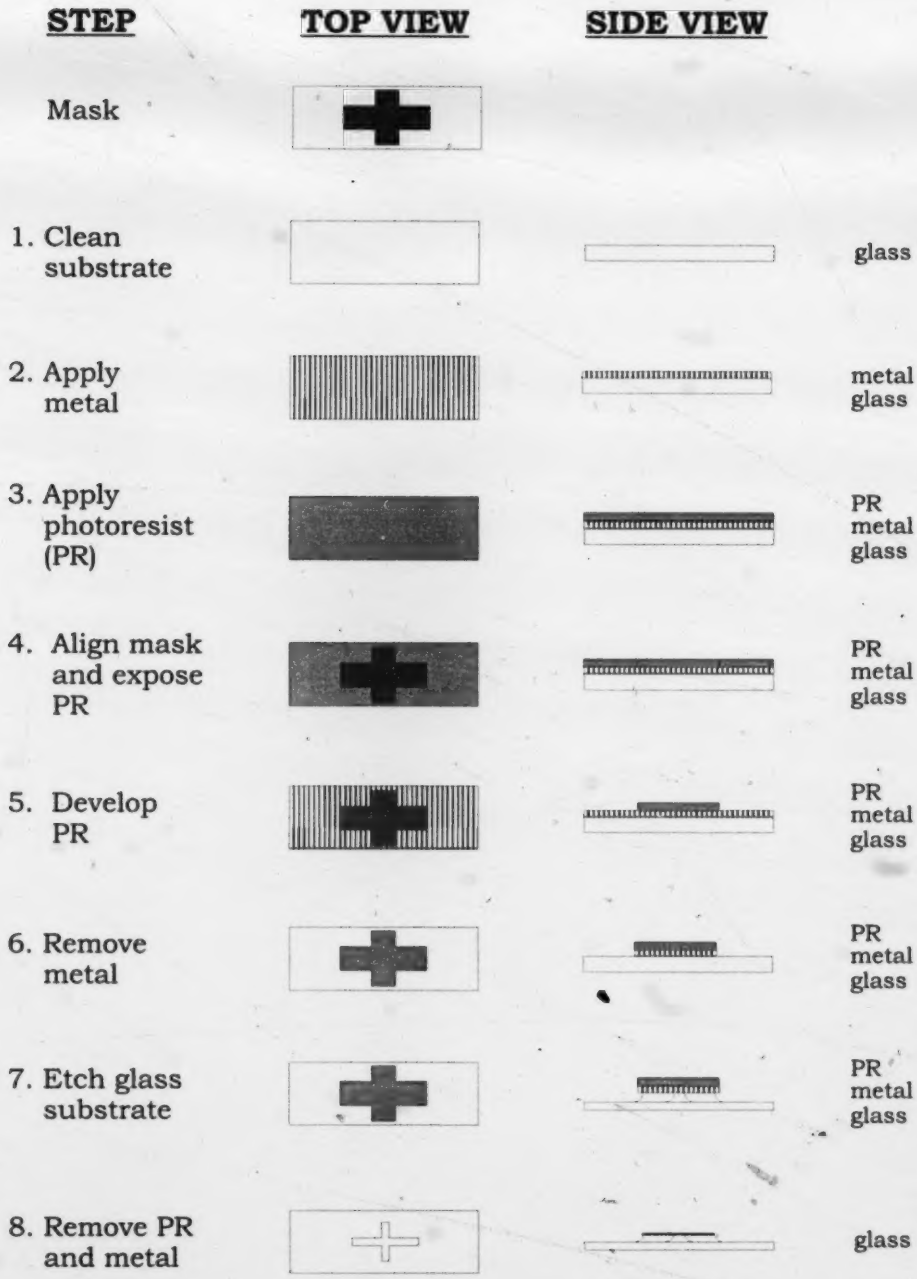


Figure 2.3. Basic Steps of Photolithography.

Figure reproduced with permission from H.J. Crabtree, 1997.

The rapid growth of the computer and electronics industries is attributed to the fabrication of miniature electronic devices, the integrated circuit, in silicon by photolithography. The silicon photolithography process is now routinely used to fabricate, or micromachine, mechanical devices such as the micro-accelerometers that deploy automobile airbags and the micro-pumps that dispense dye in ink-jet printers [3]. Recently, micromachining of glass substrates has found applications in chemistry and biotechnology such as microchips that analyze blood with electrochemical sensors [3]. Many groups are currently investigating a miniaturized total analysis system (μTAS) to perform sample handling, reactions, separation, and detection [4-6].

Our group has developed a glass micromachining process to fabricate sheath-flow cuvette detection cells for capillary electrophoresis [7]. In the sheath-flow cuvette, buffer flows between the capillaries and focuses the DNA sample stream as it exits the capillary. The laser beam is positioned under the capillary, exciting the fluorescently labeled DNA. In the 16-capillary cuvette, structures are etched on the inside of the glass cuvette to fix the capillaries in a tight linear array. The design of the cuvette is shown in Figure 2.4. The etched grooves in the glass, called capillary guides, evenly space the capillaries. The capillaries are held at the same height above the laser beam by "V" grooves, or capillary stoppers, which catch the edges of the capillary. These glass structures ensure that the fluorescent streams from each capillary are reproducibly aligned to the detection optics.

2.1.2. Detection Apparatus

Optics

Several choices were made for the design of the detection optics in order to monitor the fluorescence from all the capillaries simultaneously. Microscope objectives have high collection efficiencies compared to simple lenses. Microscope objectives are also color-corrected for 656 nm and 486 nm [8]. Because of their superior performance, microscope objectives were used to image the fluorescence from the cuvette onto an array of Gradient Index (GRIN) lenses (Figure 2.2).

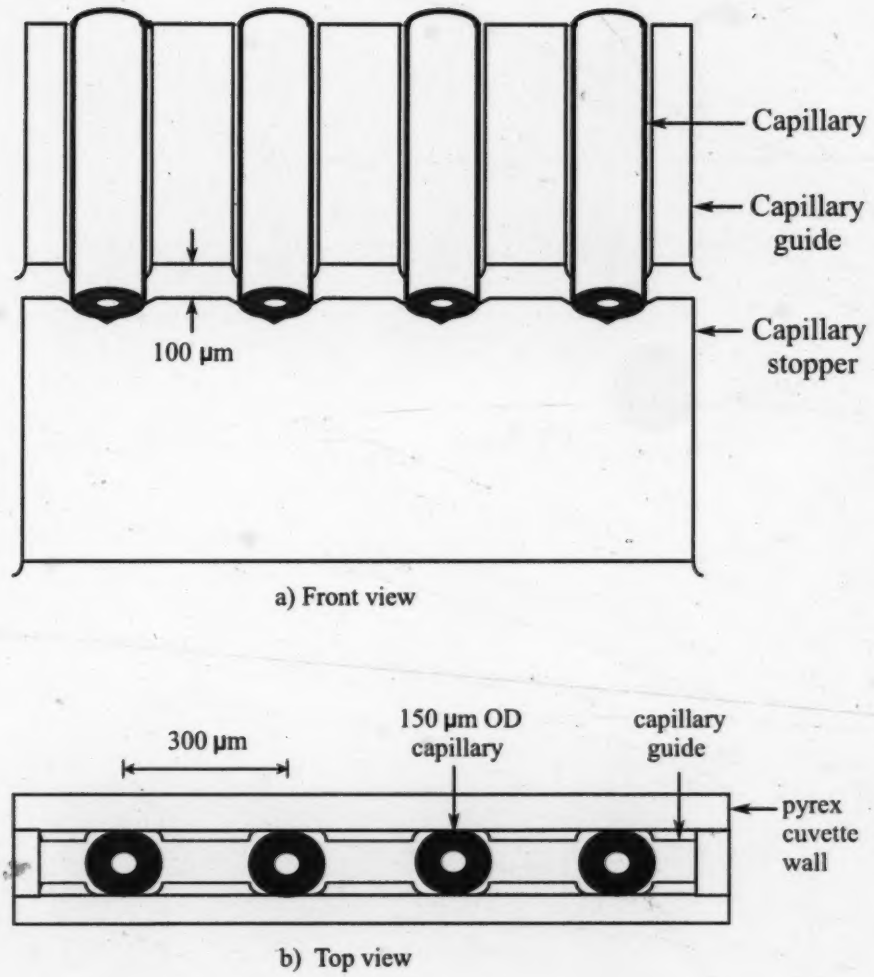


Figure 2.4. Capillary Alignment in the Micromachined Cuvette.
Capillaries are held in a linear array, spaced 300 µm apart by capillary guides. The capillary stoppers hold the capillaries at a specific height. Figure reproduced with permission from H.J. Crabtree, 1997.

GRIN lenses [9, 10] are compact, inexpensive and simple optical elements that focus the fluorescent light from the image surface to a small spot, matching the small light-sensitive area of the APD. In conventional optics, light is focused by the curvature of the lens surface (Figure 2.5.a.). The lens is composed of material of a single refractive index. The material's refractive index determines how light will bend at a surface according to Snell's Law. In GRIN lenses, the incident light is focused by a radial variation in the refractive index. The radial refractive index variation, n(r), is described by Equation 2.1:

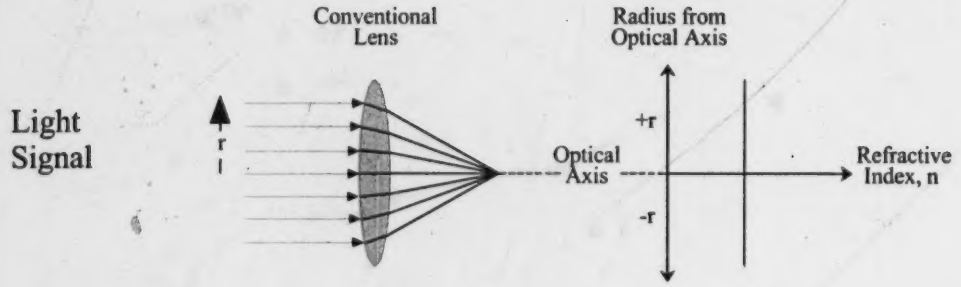
$$n(r) = n_0\left(1 - \frac{A}{2}r^2\right) \qquad (2.1)$$

where n_0 is the base refractive index at the optical axis, A is the squared gradient constant, and r is the radial position of the lens. Because the refractive index varies radially, light moves in a sinusoidal path along the optical axis. One pitch is the length of the GRIN lens needed for light of a certain wavelength to execute one cycle or one period. A GRIN lens with 0.25 pitch is able to focus light from a collimated beam to a point, or vice versa, as depicted in Figure 2.5.b.

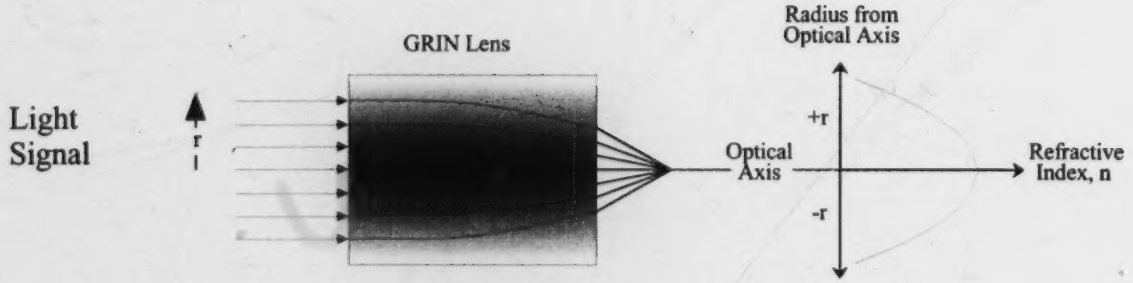
Avalanche Photodiodes

Several design considerations influenced the choice of the photodetector for the 16-capillary instrument. Photomultiplier tubes (PMT) are useful for low-light applications because they have a linear response over several orders of magnitude of light intensity and have high internal gain (10^6). However, their bulky size makes PMTs impractical and cumbersome to use in instruments that require multiple detectors. A charge-coupled device (CCD) camera can detect multiple channels in the 16-capillary instrument. However, CCD cameras are limited in the speed that the data can be read out, and most importantly, CCD cameras are expensive. The use of APD detectors was less expensive compared to the use of a CCD camera.

Solid-state devices, such as silicon APDs, are compact, robust, and have fast response. They also have high quantum efficiencies, 70-95% from 500 to 850 nm [11], where the quantum efficiency is defined as the ratio of number of current carriers generated to the number of incident photons. APDs require fiber optic coupling to



a) Conventional Lens



b) Gradient Index (GRIN) Lens

Figure 2.5. Conventional and GRIN Lenses.
In a), light is focused by the curvature of the lens. In b), light is focused by the change in refractive index of the lens.

transfer the light to the detector because of their very small light-sensitive area.

A classic description of diodes and photodiodes can be found in electronics textbooks [12]. An APD is a photodiode that generates electron-hole pairs in two stages leading to internal current gain. A schematic cross-section of a typical "reach-through" structure APD is shown in Figure 2.6 [13]. As in photodiodes, light of the appropriate energy creates electron-hole pairs in the absorption region *A*. An electric field across *A* sweeps the photogenerated (primary) electrons to the *pn* junction, also called the multiplication region *M*. The *M* region has a high electric field to accelerate the carriers and produce secondary carriers by impact ionization of the silicon. Figure 2.7 shows the current-voltage curve for an APD.

The APD temperature must be strictly regulated because the ionization rate of electron-hole pairs depends on temperature. As a result, the APD biasing voltage has a temperature coefficient; the breakdown voltage decreases as the APD is cooled. The temperature at which the APD is operated determines the dark signal, which is the signal that is produced in the absence of radiation. Thermally generated electrons create photoelectrons that generate a background current. Cooling the APDs reduces the dark signal; the dark current roughly doubles for every +10°C [14].

APDs can be operated in two different modes, Geiger or linear, depending on the biasing voltage. In Geiger mode, or photon-counting mode, the APD is biased higher than the breakdown voltage and the APD conducts a large current [11]. In this mode, a single photoelectron may trigger an avalanche pulse with gain of 10^8 [15]. In the linear mode, the APD is biased lower than breakdown voltage. In this voltage region, internal gains of 10^2 can be achieved. An external amplifier is usually required to amplify the current signal from the APD [15]. The current generated is linearly proportional to the number of incident photons.

For the 16-capillary instrument, the high sensitivity, compact size, robust nature and high quantum efficiencies made the use of APDs an interesting choice. Sub-nanosecond response times were not important in our applications. The APDs in this DNA sequencing application (C30902S, EG&G Canada, Vaudreuil, Quebec) were suited to Geiger mode because they have low dark current, which is aided by cooling to -20°C.

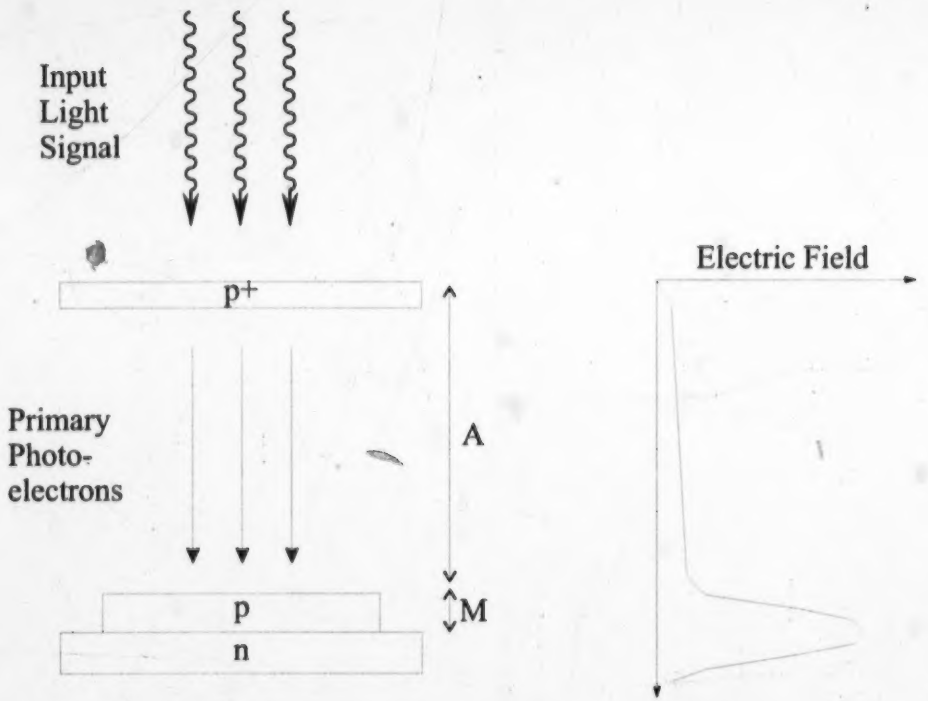
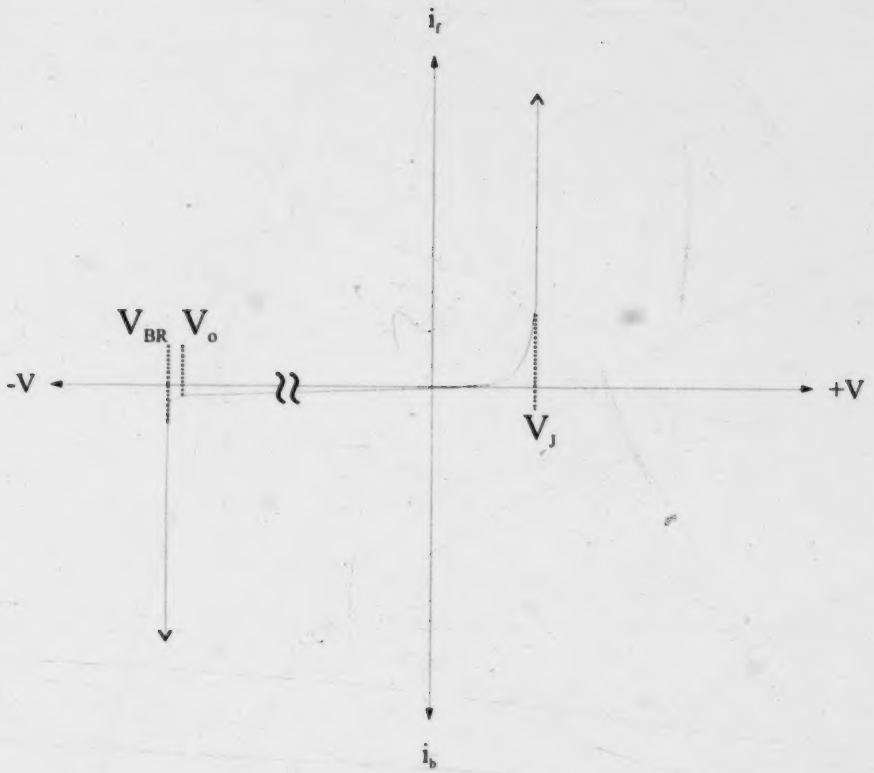


Figure 2.6. Diagram of a Reach Through Structure APD (not to scale).

Photoelectrons, generated from incident light are accelerated towards the Multiplication Region (M) by an electric field across the Absorption Region (A). The Multiplication Region (M) exhibits a high electric field to provide internal current gain by impact ionization.



V_{BR}, breakdown voltage
V_O, operating voltage (linear mode)
V_j, junction voltage
i_f, forward current
i_b, reverse current

Figure 2.7. Current-Voltage Curve for a Silicon Avlalanche Photodiode.

However for simplicity of electronic construction, the linear mode operation was chosen.

2.1.3. DNA Sequencing

Once the 16-capillary instrument was built, DNA sequencing samples were run to verify that the data from all 16 channels could be acquired simultaneously. Single-stranded (ss) M13mp18 bacteriophage [16] was used as the DNA sequencing sample control template. In addition to sequencing the control template, some fragments of wild-type (wt) M13 bacteriophage cloned into double-stranded plasmid pUC19 [16] were also sequenced to evaluate the instrumentation.

2.2. EXPERIMENTAL

2.2.1. Separation Apparatus

2.2.1.1. Micromachined Sheath-flow Cuvette

Corning Pyrex #7740 (Paragon Optical Company, Inc., Reading, PA) was used to fabricate the cuvettes. This glass transmits light from 360 nm to 1800 nm with 90% transmittance [17]. The glass substrate size was ~10 cm × ~10 cm × 1 mm. The substrate was annealed at 570°C for 2 h to remove stresses in the glass that otherwise lead to a frosty appearance when etching with HF [7].

The substrate surface was cleaned with "pirhana" (3:1 volumetric mixture of 96% sulfuric acid and 30% hydrogen peroxide) for proper adhesion of the metal and photoresist layers. Poor adhesion resulted in etching under the metal and photoresist layers and reduced the resolution of the etched features. Thin metal layers were deposited onto the cleaned substrate surface by an Ar^+ plasma sputtering system (model 4410, Perkin-Elmer, Mississauga, Ontario). Three layers of metals were deposited on the glass surface: 200 nm of chromium, 100 nm of gold and 200 nm of chromium.

The positive photoresist (HRP 506, Olin Corp., Stamford, CT) was spin-coated onto the metal-coated substrate for 5 s at 500 revolutions per minute (rpm) and 30 s at 2000 rpm (model 5100, Solitec, Santa Clara, CA). The photoresist layer was about 1.0 μm thick. The photoresist was soft-baked (30 min at 110°C) to remove photoresist solvent.

The mask (Precision Photomask PPM, St. Laurent, Quebec), with the cuvette designs, was aligned over the photoresist-covered substrate and exposed to UV light for about 5 seconds (model MA4, Karl Suss KG-GmbH & Co., Munchen-Garching, Germany). A 1:1 positive image was transferred to the photoresist layer after the exposed photoresist was developed with NaOH (5 s water at 500 rpm, 70 s NaOH at 1000 rpm, 90 s water at 1000 rpm and 30 s air dry at 1500 rpm). The substrate was hard-baked (30 min at 120°C) to harden the photoresist.

The metal layers under the developed photoresist were removed by sequentially submerging the substrate for a few seconds in Chrome-Etch, aqua regia and then finally Chrome-Etch solutions. Chrome Etch (Olin Corp.) removed the chromium layers and aqua regia (3:1 ratio of 37% HCl to 70% HNO_3) removed the gold layer. The substrate was etched with a buffered oxide etch (BOE, 10:1 volumetric ratio of 40% NH_4F to 49% HF) at room temperature. The glass etched at a rate of 1 µm/hour, which was monitored by a profilometer (model 10-02000, Tencor Instruments, Mountain View, CA). The patterns were usually etched to a depth of 25 µm. Once the desired depth was obtained, the substrate was washed, the remaining photoresist was removed with acetone and then the metal layers were removed as described above. Aspect ratios, the ratio of horizontal etch to vertical etch, of 1:1 were obtained routinely.

The individual cuvette pieces, typically 25 mm × 12 mm, were cut from the substrate using a wafer-dicing saw (Model 150, Micro Automation General Signal, Freemont, CA). A 0.25-mm thick diamond-impregnated resinoid blade spinning at 22000 rpm was used to cut the pieces. The edges were polished using a lapidary machine (Buehler Ecomet, Lake Bluff, IL) with 400, 800, 3000, 8000 and 12000 grit diamond paste. The cuvette was assembled from 4 pieces: two identically etched pieces and two spacers made from etched cover slips, about 135 µm thick. The assembly is shown in Figure 2.8. The inner edges of the cover slip were glued together as a stack and polished with lapidary equipment to reduce laser scatter. The pieces were assembled using UV-curing glue (Norland Optical Adhesive #61, Norland Products, Inc., New Brunswick, NJ), cured for 30 minutes by illumination at 366 nm and then heat-cured for 12 hours at 55°C.

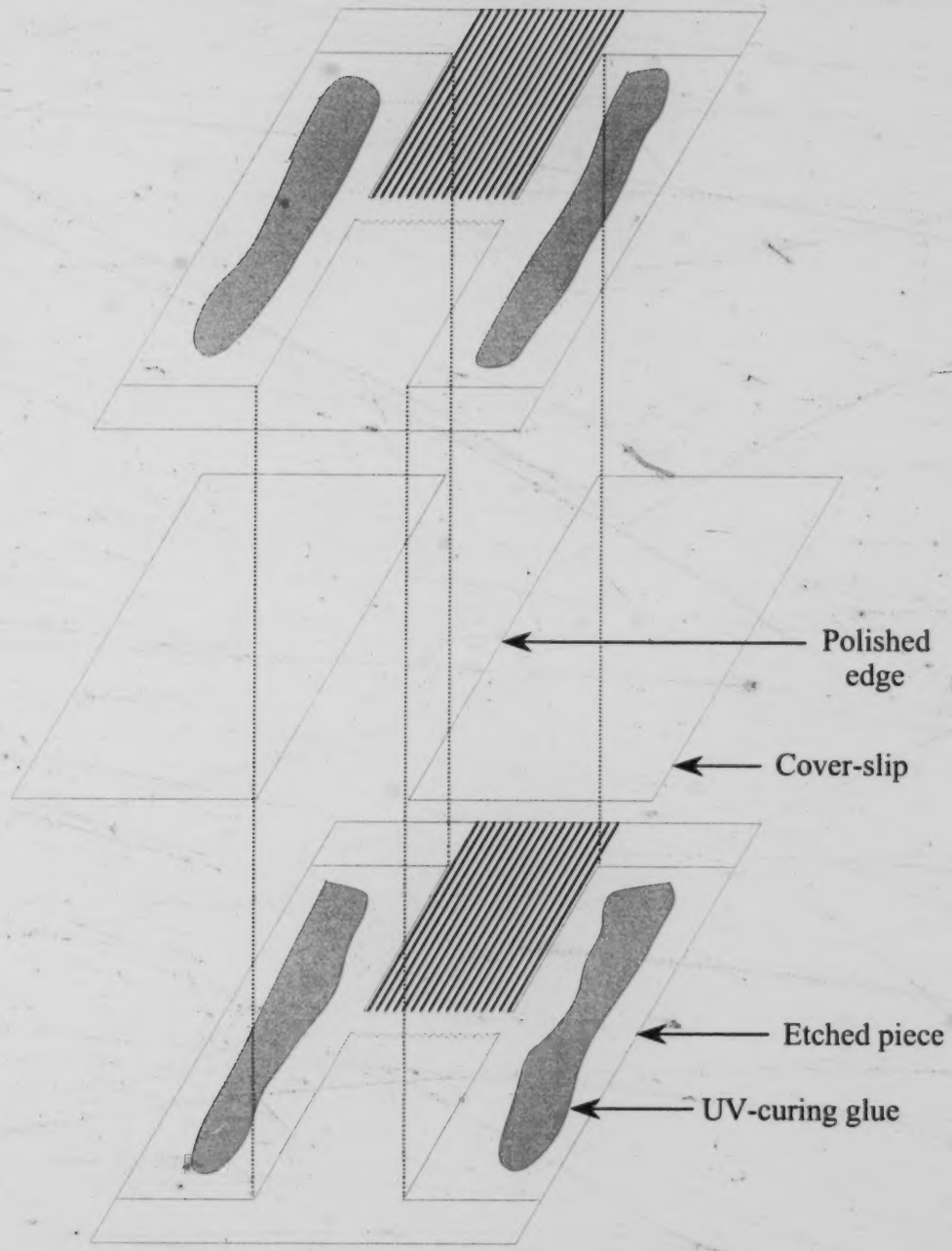


Figure 2.8. Micromachined Cuvette Assembly.
The cuvette is assembled from four pieces: two etched pieces with capillary guides and stoppers and two glass coverslips with polished edges. The pieces are glued together with UV-curing glue. Reproduced with permission from H. J. Crabtree, 1997.

Once assembled, the cuvette was cut to a final size of 21 mm × 7.2 mm using the wafer-dicing saw.

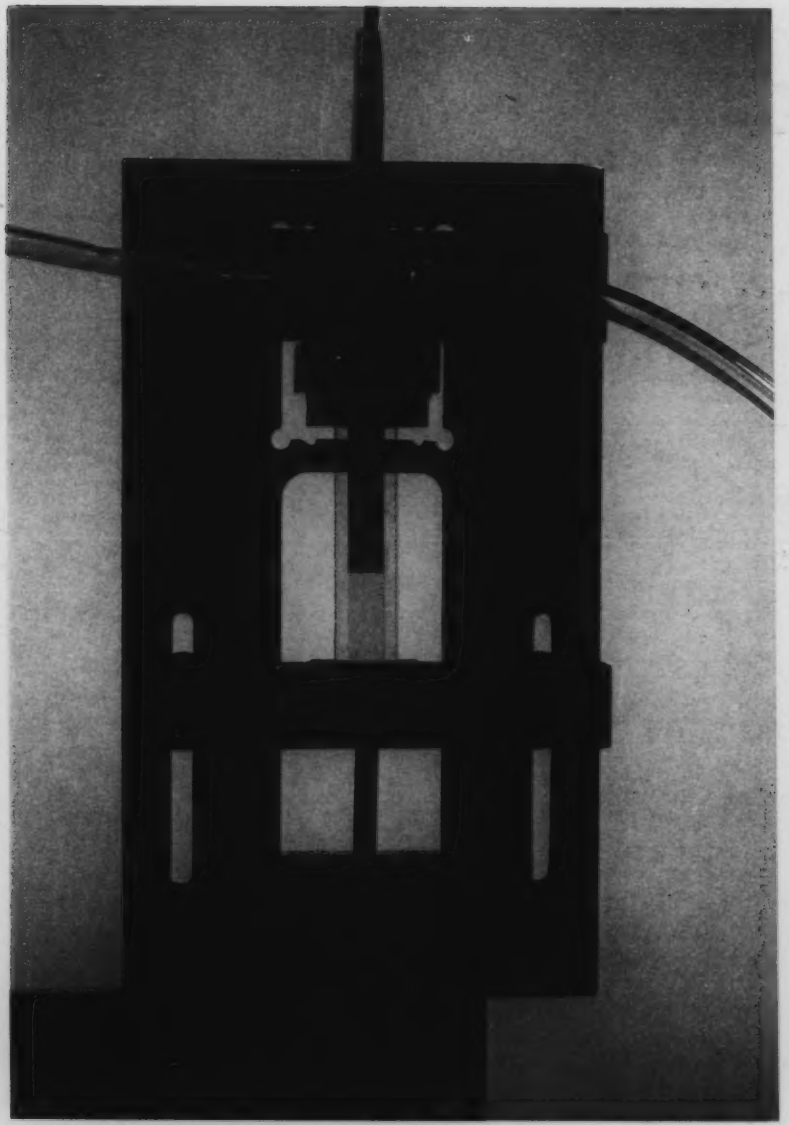
The cuvette was held in a stainless steel fixture in Figure 2.9. The detector ends of the capillaries were inserted through the top in a bundle and spread into a linear array in the cuvette. The capillaries were glued into the holder with 5-minute epoxy. The tubing on one side of the holder transported the sheath-flow buffer to the cuvette, and the tubing on the other side removed bubbles caught in the cuvette. The sheath-flow fluid exited through the tubing at the bottom of the cuvette holder. The steel holder was electrically grounded to the optical table.

2.2.1.2. Plexiglas Heating Box and Sample Injector

The first 30 cm of the capillary from the injection end was housed in a Plexiglas box (Figure 2.10). The box was 6" × 6" × 10". The box had a safety switch that disabled the high voltage power supply when the box was opened, protecting the user from electrical shock. The box was also fitted with a forced air heater (Air-Therm, World Precision Instruments, Inc., Sarasota, FL) to heat the capillaries during the run. The side-arm of the heating unit was hinged to swing up and out of the way when capillaries were replaced. The side-arm was 6" × 4" × 3". A controller regulated the temperature using a temperature sensor (Nickel 120Ω Resistive Temperature Detector-RTD) positioned approximately in the middle of the box as shown in the Figure 2.10.

Electrophoresis was driven by a high voltage (HV) power supply (CZE 1000R, Spellman, Plainview, NY). Four platinum electrodes were soldered to the high voltage cable from the power supply allowing four samples to be injected at a time. A capillary end and a high voltage platinum electrode were placed in a vial containing the DNA sample. A negative potential was applied, forcing the negatively charged DNA fragments onto the capillary (Section 2.2.3.1). The sample vial was then replaced with a vial containing buffer of the same composition as the gel buffer, and electrophoresis continued.

Figure 2.9. The 16-Capillary Cuvette Mounted in Holder.
The micromachined cuvette is mounted in a stainless steel holder. The holder is electrically grounded. The capillaries enter the top of the holder as a bundle and become a linear array in the cuvette. Buffer flows in the left tube and flows to waste out the bottom of the holder. The right tube is used to purge bubbles from the holder.



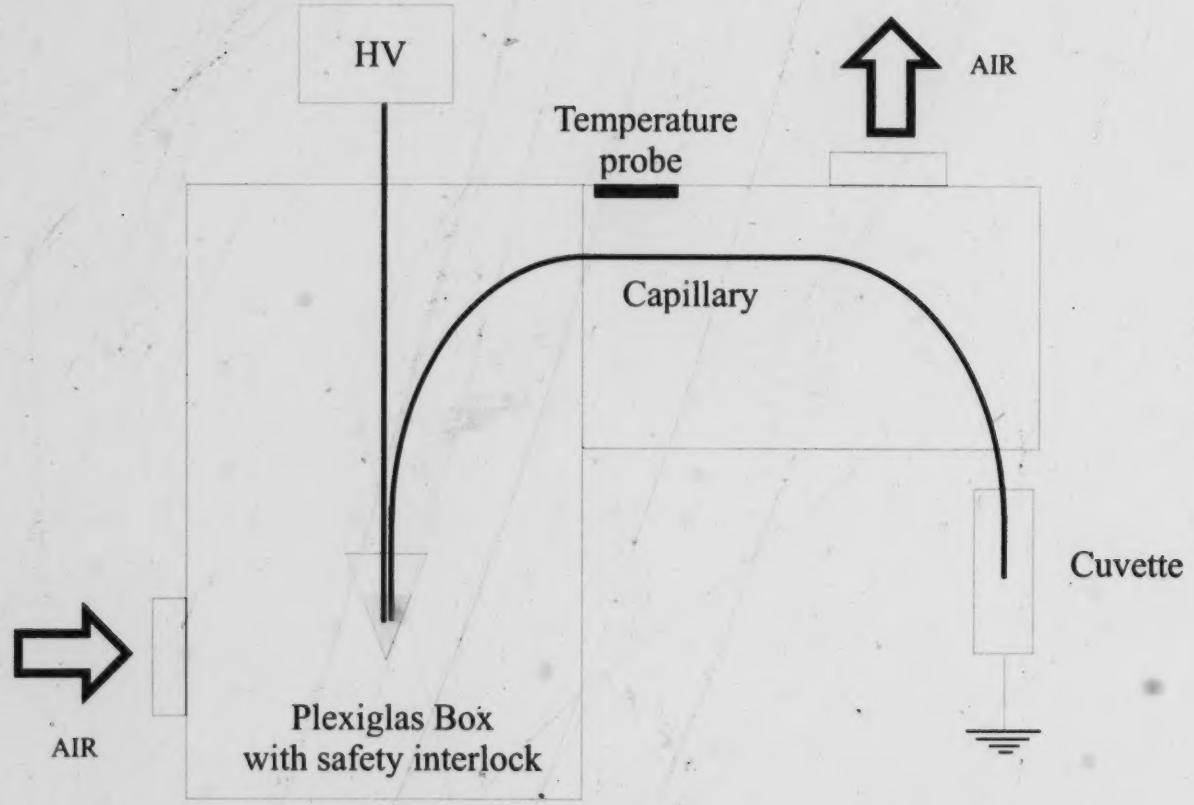


Figure 2.10. Plexiglas Capillary Box with High Voltage and Forced Air Heating (not to scale).
The Plexiglas box has a safety interlock that disables the high voltage if the box is opened. The main Plexiglas box on the left was 6" long, 6" wide and 10" tall. The side-arm was 6" long, 4" wide and 3" tall. Heated air flows in the main box and out the arm. The temperature probe is placed on the top of the box.

2.2.2. Detector Apparatus

2.2.2.1. Optics

The DNA sequencer was built for 4-color DNA sequencing chemistry, where each of the four terminators A, C, G or T are associated with a different fluorescent dye. The dyes used are: FAM (fluorescein), JOE (a fluorescein derivative), TAMRA (tetramethyl rhodamine), and ROX (Texas Red). The dye-labeled primers (PE Applied Biosystems) excitation and emission maxima and the bandpass filters are listed in Table 2.1.

Table 2.1. Fluorescent Dyes and Bandpass Filters

DNA Base	Fluorescent Dye Primer	λabs, max (nm)	λem, max (nm)	Laser Excitation (nm)	Bandpass filter (nm)
A	FAM	494	521	488 (Ar^+)	540DF10
C	JOE	525	555	488 (Ar^+)	560DF10
G	TAMRA	555	580	532 (Nd:YAG)	580DF10
T	ROX	585	605	532 (Nd:YAG)	610DF10

In Figure 2.2, the 20 mW 488 nm Ar^+ laser (Cyonics 2213-755L, San Jose, CA) and 10 mW 532 nm frequency-doubled Nd:YAG laser (Coherent 532-10, Santa Clara, CA) beams were chopped by a chopper wheel spinning at a frequency of 2 Hz. A mirror and a dichroic beamsplitter (505DRLP, Omega Optical Inc., Brattleboro, VT) recombined the chopped laser beams. The laser beams were focused onto the cuvette by 1 × microscope objective (Melles Griot, Nepean, Ontario). The laser beams excited fluorescently labeled DNA as it exited the capillary. The fluorescence from the DNA samples was collected by an 18 ×, 0.45 NA microscope objective with a working distance of 4.2 mm (Melles Griot). The four DNA bases were distinguished by spectrally filtering the fluorescent signals with a filter wheel consisting of four 25.4 mm diameter bandpass filters: 540DF10, 560DF10, 580DF10 and 610DF10 (Omega Optical Inc.). The filter wheel was positioned 1.4 cm behind the collection microscope objective and the filter wheel was synchronized to the chopper wheel.

The fluorescent spots from the 16 capillaries were imaged onto 16 GRIN lenses (SELFOC® Fiber Collimators, FCM-00F-050-0.63, Nippon Sheet Glass Co., Tokyo,

Japan); the fluorescent spot from one capillary was collected by one GRIN lens. The GRIN lenses coupled the fluorescence signal to fiber optics (50 μm core, 125 μm cladding) that were "pigtailed" to APDs. The FC pigtail connector is like a BNC connector used in electronics; the fiber optic cables can be repeatedly connected and disconnected. A pigtail from the GRIN lens and a pigtail from the APD were mated at a FC (face contact) connector (Radiant Communications Corp., South Plainfield, NJ).

The FC connector allowed back-illumination of light through the GRIN lenses to align each lens to its fluorescent spot beneath the capillary. Back-illumination would not be possible if the GRIN lenses were connected to APDs with one continuous fiber optic. GRIN lenses were aligned to the fluorescent spots in the following manner. Gel-filled capillaries were run with aligning solutions of FAM- and ROX-labeled primers. White light illuminated the FC connector and was transmitted through the GRIN lenses and the filter wheel. The sheath-flow was adjusted so the fluorescent spots matched the size of the GRIN lens spots. A fluorescent spot in the cuvette that was 50 μm wide, when magnified 18 times, resulted in a 0.9 mm diameter image. The size of this image matched the 1-mm diameter GRIN lens. When the fluorescent spots were sharply imaged in the viewing microscope, the GRIN lenses were moved so that they also were in focus and their images superimposed on the fluorescent spots. The GRIN lens images in the cuvette are shown in Figure 2.11. The GRIN lenses were mounted in rectangular aluminum holders painted flat black. They are adjustable and fixed in positions by bolts.

All fiber optic-GRIN lens and pigtail connections were manufactured in-house according to the fiber optic polishing kit instructions (#TK-8, Radiant Communications Corp.).

2.2.2.2. Avalanche Photodiode Cooling

The APDs were stacked in a 4 × 4 array in an aluminum U-block. A single column of APDs is shown in Figure 2.12. The diameter of the APDs mounting holes was about 5.4 mm. The U-block was mounted on a thermoelectric Peltier device (CP-1.4-71-10L, Melcor, Materials Electronics Products Corp., Trenton, NJ). The Peltier device cooled the APDs to -20.0°C when driven with a power supply (Kepco MSK 10-10M,

Figure 2.11. The GRIN Lens Images in the Cuvette.

The green GRIN lens images are positioned directly under the tips of the capillaries. The light illuminates the fiber optics and it is transmitted through the fiber optics, the GRIN lenses and the FAM (540DF10) bandpass filter. The GRIN lens image is positioned to match the fluorescently labeled DNA as it exits the capillary. Only 14 GRIN lenses are visible because of the combination of microscope objectives clip the marginal rays for the photo.

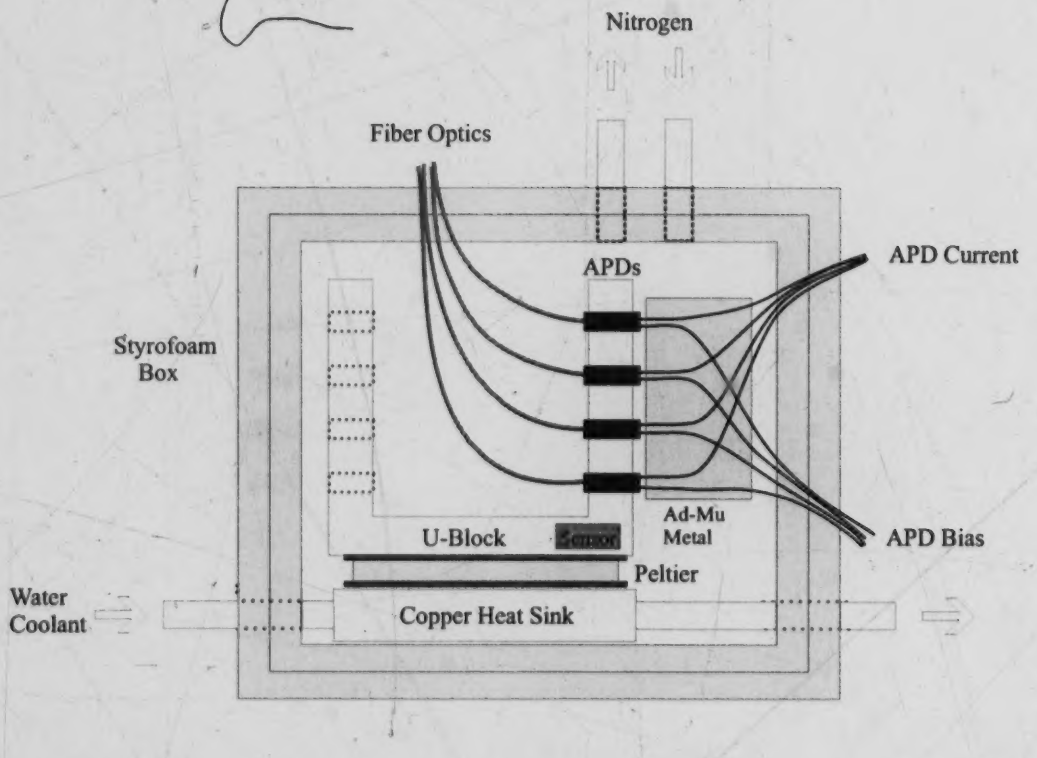


Figure 2.12. Styrofoam APD box.

The box is composed of two layers of 1" Styrofoam. The box is cooled to -20°C by a Peltier device. Heat from the Peltier is removed by pumping water through a copper heat sink. Nitrogen gas flows through the box to reduce the frost. The APDs are mounted in an aluminum U-block. APD leads are encircled with a piece of Ad-Mu metal foil for electromagnetic shielding. Fiber optic cables enter the top of the box.

Flushing, NY). The power supply was regulated by a temperature controller (μPC, ERO Electronic S.p.A., Milan, Italy) that switched between 0 A and 3.9 A, the maximum current setting for the Peltier. Mounted onto the U-block was a platinum 100 Ω RTD temperature sensor, part of the feedback loop with the temperature controller. For a precision of ±0.1°C, the settings of the temperature controller were set point = -20.0°C and proportional band = 2.0%. As the one face of the Peltier cooled the APDs, the other face heated up. The hot face was placed on top of a copper heat sink. A 30% ethylene glycol/water mixture cooled to 0°C (water bath-MGW Lauda-Brinkman RC20 and B2R Thermostat-Messgerate-Werk Lauda, West Germany) was pumped through the copper heat sink in a serpentine pattern.

The U-block, Peltier device and copper heat sink were placed in a Styrofoam box. The inside measurements of the box were 3" × 3" × 3". The box was constructed out of two layers of 1" thick Styrofoam creating a box within a box. The sides of the box were glued together with white RTV silicone (GE Silicones, Pickering, Ontario); however the inner box was not glued to the outer box. This box construction left an insulating layer of air to prevent heat transfer through the silicone joints. Nitrogen gas was circulated through the box to prevent frost formation when cooling to -20°C. The small positive pressure of nitrogen leaking out of box prevented moist air from entering. A small batch of Drierite (W.A. Hammond Drierite, Xenia, OH) was kept in the inner box to absorb moisture and also to act as a qualitative moisture gauge.

2.2.2.3. Avalanche Photodiode Electronic Circuitry

A schematic of the APD electronics is shown in Figure 2.13. The APDs were reverse biased at around -200 V by individual HV power supplies (model 522-1, Analog Modules, Inc., Longwood, FL). Light incident on an APD was converted into a current usually in the nanoampere range during DNA sequencing. The bucking current, opposing the current from the APD, offset the dark current levels before the external amplification. The current was converted to voltage and externally amplified with a gain of 10^9 V/A by a high-gain transimpedance amplifier (model 341, Analog Modules, Inc.). These specially modified amplifiers have low noise, high gain and low drift. Replacing the

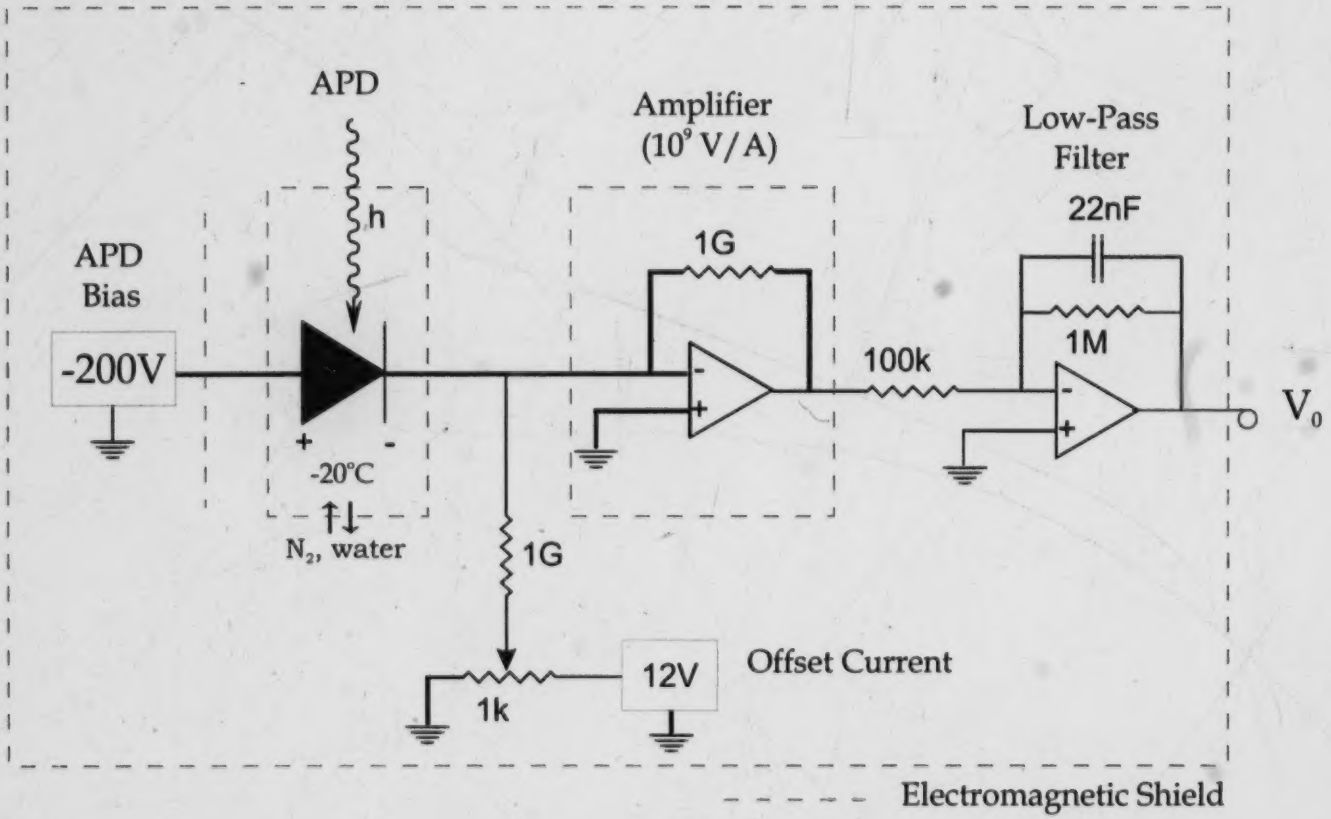


Figure 2.13. The Avalanche Photodiodes Electronics.

The APDs are reverse biased at -200 V and cooled to -20°C. Light travels through a fiber optic to the APD. The light is converted to current in the APD. An opposing current offsets the dark current in APDs. The current is amplified and converted to voltage with gain (10^9V/A). The signal is low-pass filtered and readout to the computer.

resistor in the feedback loop to 1 GΩ changed the gain in the amplifiers. To ensure conduction through the very high value resistor, the resistor surfaces were cleaned with soap and ethanol. Low-pass active filters (OP07 operational amplifiers) with an 7 Hz cutoff filtered the amplified signal. The low pass filters had a gain of 10. The filtered signal was collected by an analog-to-digital data acquisition board (NB M10-16X, National Instruments Corp., Austin, TX), digitized (16 bit) and recorded in Labview Version 3.1.1 (National Instruments Corp.) on a Macintosh computer (Apple Power Macintosh 7100/66, Apple Computers Inc. Cupertino, CA).

2.2.2.4. Avalanche Photodiode Biasing Voltage

Each APD was biased by its own HV power supply (model 522-1, Analog Modules, Inc.). The HV power supply was chosen for its low output noise and low ripple, ≤1 mV at 1 mA at -250 V output.

The APDs were cooled to -20°C and their biasing voltages were chosen to maximize the signal-to-noise ratio. A green LED was coupled to a GRIN lens connected to the APD as shown in Figure 2.14. The APD output was amplified with a gain of 10^9V/A, low-pass filtered and collected in Labview Version 3.1.1. The signal and noise in the signal were collected at various APD biases, from the breakdown voltage to 10 V beneath the breakdown voltage. The APD bias with the maximum signal-to-noise ratio was chosen for the operating voltage at -20°C.

2.2.2.5. Chopper/Filter Wheel and Data Acquisition (DAQ)

The chopper and the filter wheel served two purposes: to spectrally discriminate the four fluorescent dyes and to trigger the data acquisition.

Spectral discrimination was achieved by the following procedure. The chopper wheel was synchronized to the filter wheel by a single controller driving the two stepper motors. The two wheels rotated at a frequency of 2 Hz. The 488 nm laser produced water Raman scatter around 580 nm. To reduce the interference from the Raman scatter, the Ar^+ laser illuminated the sample when the FAM (540 nm) and JOE (560 nm) bandpass filters were positioned behind the microscope objective. To eliminate Rayleigh

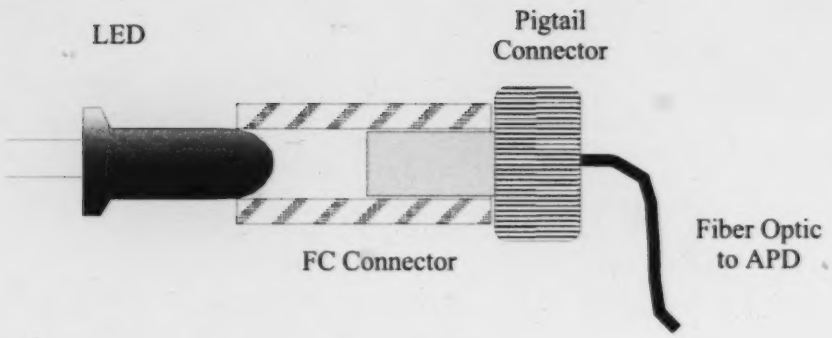


Figure 2.14. The LED Coupling to FC Connector
To set the APD bias, a green LED was glued to an FC connector. A fiber optic pigtail connector transmitted the LED light to the APD.

interference of Nd:YAG laser (532 nm) in the FAM bandpass filter, the Nd:YAG laser illuminated the sample when the TAMRA (580nm) and ROX (610 nm) bandpass filters were in position [18].

A detailed description of the wheel controller and data acquisition can be found in the theses of Lewis [19] and Crabtree [7]. Very briefly, an interrupter device detected notches cut into the filter and chopper wheels. The notches trigger the sequential data acquisition of the fluorescent signal through each of the four bandpass filter. The notches also served to synchronize the chopper and filter wheels; the wheels would autoalign if they fell out of phase.

2.2.3. DNA Sequencing

2.2.3.1. Single-Stranded Template Sequencing

Sample Preparation

The cycle sequencing reactions were prepared using an Amersham RPN 2536 kit with ThermoSequenase DNA polymerase (Amersham Life Sciences, Buckinghamshire, England) and dye-labeled primers (Kit 401487, PE Applied Biosystems Inc., Foster City, CA). The four termination reactions were prepared in four separate tubes. The reagent mix contained the dNTPs, the enzyme, buffer and the appropriate ddNTP (see Appendix I). A total of 1.4 µg of ssM13mp18 (Amersham Life Sciences) template was used and the sample was prepared as shown in Table 2.2. The final volumes were made up with water.

Table 2.2. M13mp18 Cycle Sequencing ss Template DNA

	A-termination	C-termination	G-termination	T-termination
Reaction Ratio A:C:G:T	2	2	2	1
Template	0.40 µg	0.40 µg	0.40 µg	0.20 µg
M13 Forward Dye-Primer	1.6 pmol (JOE)	1.6 pmol (FAM)	1.6 pmol (TAMRA)	0.8 pmol (ROX)
Reagent Mix	8 µL (ddATP)	8 µL (ddCTP)	8 µL (ddGTP)	4 µL (ddTTP)
Final Volume	32 µL	32 µL	32 µL	16 µL

Cycle sequencing was performed for 30 cycles of 30 s at 95°C and 30 s at 55°C. The four tubes were combined in a Microcon 30 tube (Amicon, Beverly, MA). The tubes

were spun at 11000 rpm for 4-5 min in a microcentrifuge. The membrane was washed with 200 µL of water. The tubes were spun again for 4-5 min at 11000 rpm. The tubes were inverted into a new tube and the cleaned DNA in a small volume of water was spun out of membrane after 4-5 min at 6000 rpm.

Gel Preparation

Capillaries were 50 µm ID, 150 µm OD, and typically 40 cm long (Polymicro Technologies, Phoenix, AZ). The capillaries were cut in batches of ~200 using the wafer-dicing saw with a 0.25-mm blade.

The inner walls of the capillaries were coated using Hjerten's method to covalently bind the polyacrylamide to the walls and prevent electroosmosis, movement of the bulk solution resulting from the charged surface of capillary wall [20]. A solution of 20 µl of γ-methacryloxypropyltrimethoxysilane (Sigma, St. Louis, MO) in 500 µL of water and 500 µL glacial acetic acid was pulled by vacuum through the capillaries for approximately 30 min.

To polymerize the gel in the capillaries, the injection ends were placed in a 50-ml 3-neck flask. The acrylamide solution was 4%T 0%C in 7 M urea, 0.4 g acrylamide and 4.2 g urea diluted to 10 ml with 1×TBE buffer (89 mM Tris, 89 mM boric acid and 2 mM ethylenediaminetetraacetic acid (EDTA), pH=8.0). The solution was passed through a 0.22 µm filter and added to the flask. The solution was degassed with a vacuum pump for 10 minutes. The polymerization of the acrylamide was initiated under an argon blanket by adding 10 µL of 10% w/v ammonium persulfate and 5 µL of TEMED (N,N,N',N'-tetramethylethylenediamine). The solution was immediately drawn through the capillaries by vacuum, with the detector ends of the capillaries under water. The water helped to visualize the higher refractive index gel solution and to prevent urea precipitation, which occurred as the solution evaporated in air. The vacuum was removed and the gel was left to polymerize overnight. Capillaries were visually inspected for bubbles before insertion into cuvette.

The gels were pre-run with 400 nM solution of FAM- and ROX-labeled primers in 1×TBE for 30 min at 150 V/cm until the capillary current stabilized. The size and

position of the fluorescent spots of the solution were matched to the GRIN lens spots. The aligning solution was then migrated from the capillary by electrophoresis of 1×TBE for 15 min before sample injection.

Sample Injection and Run Conditions

All 16 capillaries were placed into a single sample vial. The cycle sequencing reaction was injected on all 16 capillaries for 30 s at 100 V/cm. The sample vial was replaced with 1×TBE running buffer. The electrophoresis was continued at 200 V/cm at room temperature. The first generation version of the filter/chopper wheels, without the synchronization circuitry, was used for these sequencing runs.

2.2.3.2. Cloned Wild-Type M13 Sequencing

Sample Preparation

The replicative form of wild-type (wt) M13 (6.4 kbp, Gibco BRL, Gaithersburg, MD) bacteriophage was cloned by inserting randomly sheared, blunt-end fragments into pUC19 plasmid. The brief protocol follows. Wt M13 was sonicated at 0.45 Hz for 5 s. The sheared M13 DNA was run on an agarose gel and 0.6-1 kbp DNA fragments of DNA were selected by cutting the gel slice. The DNA was extracted from the gel slice, cleaned using the GeneClean kit (Bio/Can Scientific, Mississauga, Ontario) and the ends were polished with Klenow enzyme. The wt M13 fragments were ligated into pUC19 vector (2.7 kbp, Gibco BRL), which was blunt ended with *Hinc*II and treated with shrimp alkaline phosphatase (SAP, Amersham Life Sciences). The recombinant vector was transformed into *E. coli* strain DH5α. The colonies were screened for inserts using α-complementation (blue/white colony selection) [21]. The recombinant vector was screened for the insert DNA by amplifying across the multiple cloning site by PCR (AB PCR, see Section 5.2.2 for more details).

The double-stranded pUC19 plasmids, 3.2 to 3.7 kbp long, were used as sequencing templates. The recombinant DNA was isolated from bacteria using the alkaline lysis miniprep protocol [21]. The double-stranded plasmid DNA were cut with *Sca*I to linearize the templates for sequencing reactions.

Approximately 2.2 μg total of double stranded plasmid template was required for sequencing reactions for the cycle sequencing recipe shown in Table 2.3.

Table 2.3. Wild-Type M13 Cycle Sequencing ds Template DNA

	A-termination	C-termination	G-termination	T-termination
Reaction Ratio A:C:G:T	1	1	1.5	2
Template	0.40 μg	0.40 μg	0.60 μg	0.80 μg
M13 Forward or Reverse Dye-Primer	0.8 pmol (JOE)	0.8 pmol (FAM)	1.2 pmol (TAMRA)	1.6 pmol (ROX)
Reagent Mix	4 μL (ddATP)	4 μL (ddCTP)	6 μL (ddGTP)	8 μL (ddTTP)
Final Volume	16 μL	16 μL	24 μL	32 μL

The reactions were thermal cycled and washed as described in Section 2.2.3.1.

Gel Preparation

The gels were prepared as in Section 2.2.3.1 with two changes. The silanizing solution of 20 μL of γ-methacryloxypropyltrimethoxysilane in 1 mL of 95% ethanol was drawn through the capillaries for 1 h. The gels were 6%T 0%C polyacrylamide in 7 M urea and 1×TBE.

Sample Injection and Run Conditions

After ensuring alignment of the capillaries to the GRIN lenses, the Plexiglas box was heated to 35 or 40°C. One sample was injected onto two capillaries simultaneously for 30 s at 100 V/cm. The sample vial was replaced with a vial containing polymerized gel in 1×TBE. The electrophoresis was continued at 150 V/cm at 35 or 40°C. The filter/chopper wheels setup with the synchronization circuitry constructed by D.Lewis [19] was used for these sequencing runs.

2.3. RESULTS AND DISCUSSION

Figure 2.15 shows a photo of the first generation 16-capillary sequencer in action. The lasers beams are chopped by the chopper wheel, recombined by the mirror and beamsplitter and focused onto the cuvette. Figure 2.16 shows a more detailed view of the

first generation 16-capillary sequencer. In the center of the photo, the orange fiber optic cables are connected to small black squares (GRIN lens holders); five holders are mounted on an aluminum stage. To the left of the GRIN lenses is a piece of black cardboard, part of the box covering the detection optics, and the collection microscope objective that images the fluorescent spots in the cuvette. Behind the GRIN lenses (top-center) is the blue Styrofoam box that houses the cooled APDs. The black cables from the box are fiber optic cables that are joined by FC connectors to the orange fiber optic cables.

2.3.1. Separation Apparatus

2.3.1.1. Micromachined Sheath-flow Cuvette

The design and fabrication of the cuvettes were developed by Crabtree [7]. Crabtree developed the entire photolithography procedure to etch the cuvette features, the design of the cuvette with capillary guides and "V" stoppers and the assembly of the etched pieces and cover slips into a cuvette. I helped design and construct the final version of the cuvette.

In the early cuvette designs, the capillaries were spaced 150 μm apart for 150 μm capillaries. Despite the capillary guides and stoppers, the positions of the fluorescent spots were irregular (top figure, Figure 2.17). The capillaries were numbered 1 to 16 as seen in the viewing microscope objective from left to right. Capillary 1 was the first capillary the laser beam illuminated. The capillaries were not held tightly by the capillary guides and the stoppers; the capillaries were able to move laterally by approximately ±5 μm. As a result, some capillaries pinched together, resulting in a sheath-flow imbalance between the capillaries, merging the fluorescent spots.

The inter-capillary spacing was designed to fit as many capillaries as possible in the smallest linear array. This "close-packing" would enable our collection microscope objective to collect the fluorescent spots from all the capillaries simultaneously. The

Figure 2.15. The 16-Capillary DNA Sequencer.

The blue argon ion and the green Nd:YAG laser beams originate from the back-center of the photo. The chopper wheel at the right chops the laser beams. The beams are recombined by the mirror and beamsplitter at the bottom center and they are focused onto the cuvette. Behind the cuvette is a black box, which holds the detection optics: the collection microscope objective, the filter wheel and the GRIN lenses. The left of the photo shows the Plexiglas box, which houses the capillaries.

Figure 2.16. The First Generation 16-Capillary Sequencer.
In the center of the photo, the orange fiber optic cables are connect to small black squares (the GRIN lens holders). Five holders are mounted on an aluminum stage. To the left of the GRIN lenses is a piece of black cardboard and the collection objective that images the fluorescent spots in the cuvette. Behind the GRIN lenses (top center) is the blue Styrofoam box that houses the APDs. The black cables are the fiber optic cables that connect to the orange fiber optic cables with FC connectors.

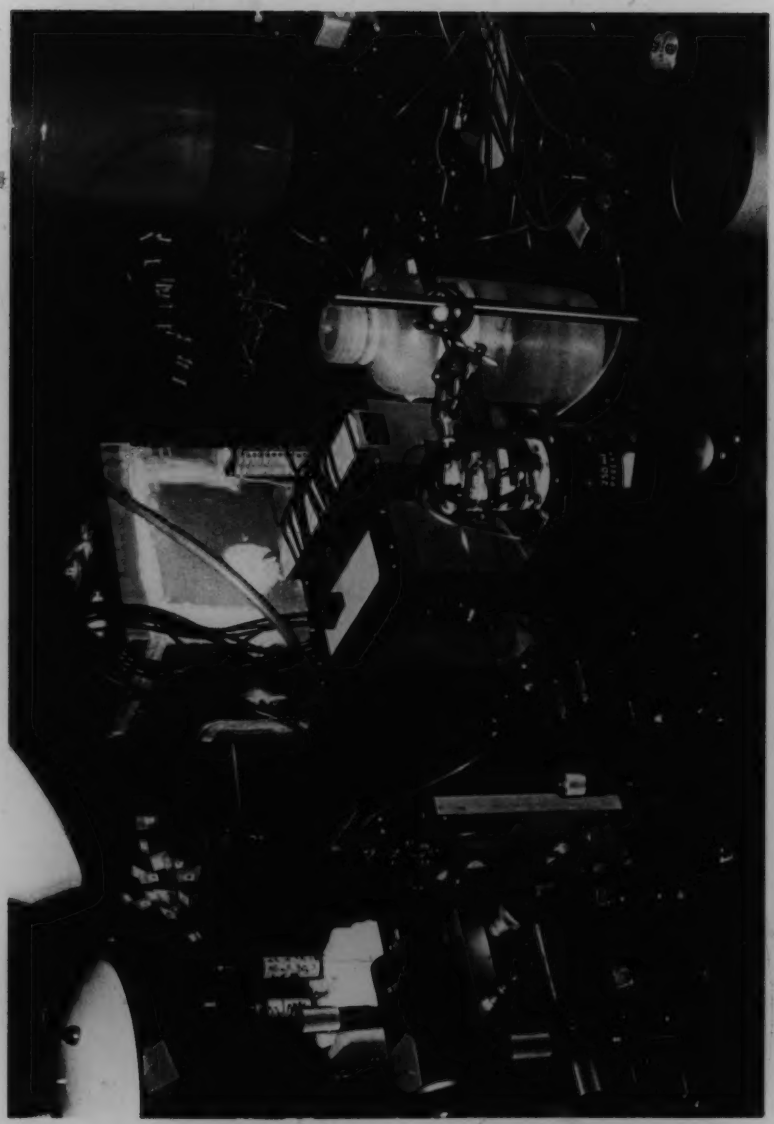
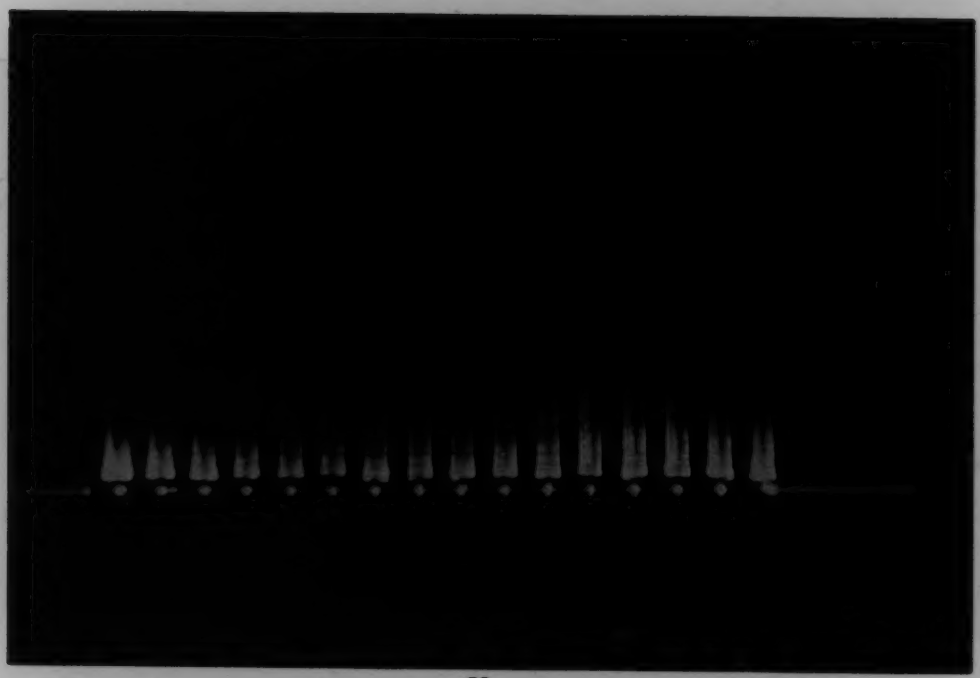


Figure 2.17. The Fluorescent Spots in the Micromachined Cuvette.
The capillaries are numbered 1 to 16 from left to right. In the top photo, the 150 μm OD capillaries are spaced 150 μm apart. Capillary 4 and 5 are pinched together resulting in a sheath-flow imbalance. As a result, the fluorescent spots are almost merged and irregular. In the bottom photo, the inter-capillary distance is increased to 300 μm and the fluorescent spots are evenly spaced and reproducible.



cuvette array width was also minimized to reduce the laser beam divergence. The capillary outer diameter determined the spacing of the capillaries in the earlier cuvette designs. The width of 16 capillaries spaced 150 μm apart was 2.4 mm. By increasing the spacing between the capillaries, small movements of the capillary did not significantly disturb the sheath flow. An inter-capillary spacing of 300 μm was constructed using a 32-capillary cuvette with a capillary spacing of 150 μm. By placing 16 capillaries in alternate capillary guides, a larger volume of sheath buffer flowed between the capillaries. Reproducibly spaced fluorescent spots were obtained with balanced sheath flow on each side of the capillary (bottom figure, Figure 2.17). The width of this 16-capillary cuvette was 4.8 mm.

The divergence of the laser beam over 4.8 mm was not significant. The Gaussian laser beam spot size, ω, a distance z from the focus is given by Equation 2.2:

$$\omega = \omega_0 \left[1 + \left(\frac{z}{z_c} \right) \right]^{1/2} \qquad (2.2)$$

where ω_0 is the beam waist spot size at z=0, $z_c = \pi n \omega_0^2 / \lambda$ is the confocal distance, n is the refractive index, λ is the wavelength of the laser [22]. From Equation 2.2, the beam is relatively collimated over a distance of $z=2z_c$. The confocal distance was equal to half the width of the cuvette, 2.4 mm for a 4.8 mm long capillary array. In the center of the cuvette, the 488 nm Ar^+ laser beam was focused to a beam waist radius, ω_0, of 20 μm. At the peripheral capillaries, the beam was calculated to expand to a 27 μm spot.

Replacing the capillaries in the cuvette was simple and fast, requiring 15 min for 16 capillaries.

2.3.1.2. Plexiglas Heating Box and Sample Injector

The Plexiglas box was a poor device to heat capillaries. The temperature was difficult to stabilize because air has a low specific heat capacity. The air heating resulted in a temperature oscillation varying from ±0.2°C to ±1.5°C around the set point temperature, 35-40°C. This oscillation was measured with a thermocouple (Digi-Sense Thermometer, Cole-Parmer Instrument Company, Vernon Hills, IL). The temperature oscillation was not noticed during the system construction because it occurred faster than

the response time of the Nickel RTD probe used with the heater. In Chapter 3, a new capillary heater made from copper is described.

Four samples were injected at a time, using one platinum electrode per sample. Four injection steps were required to load samples on all 16 capillaries; 30-40 min were required to inject all 16 samples. Obviously sequential injections of samples was tedious and impractical for large numbers of capillaries. A simple and fast method to inject all 16 sample simultaneously is presented in Chapter 3.

2.3.2. Detector Apparatus

2.3.2.1. Optics

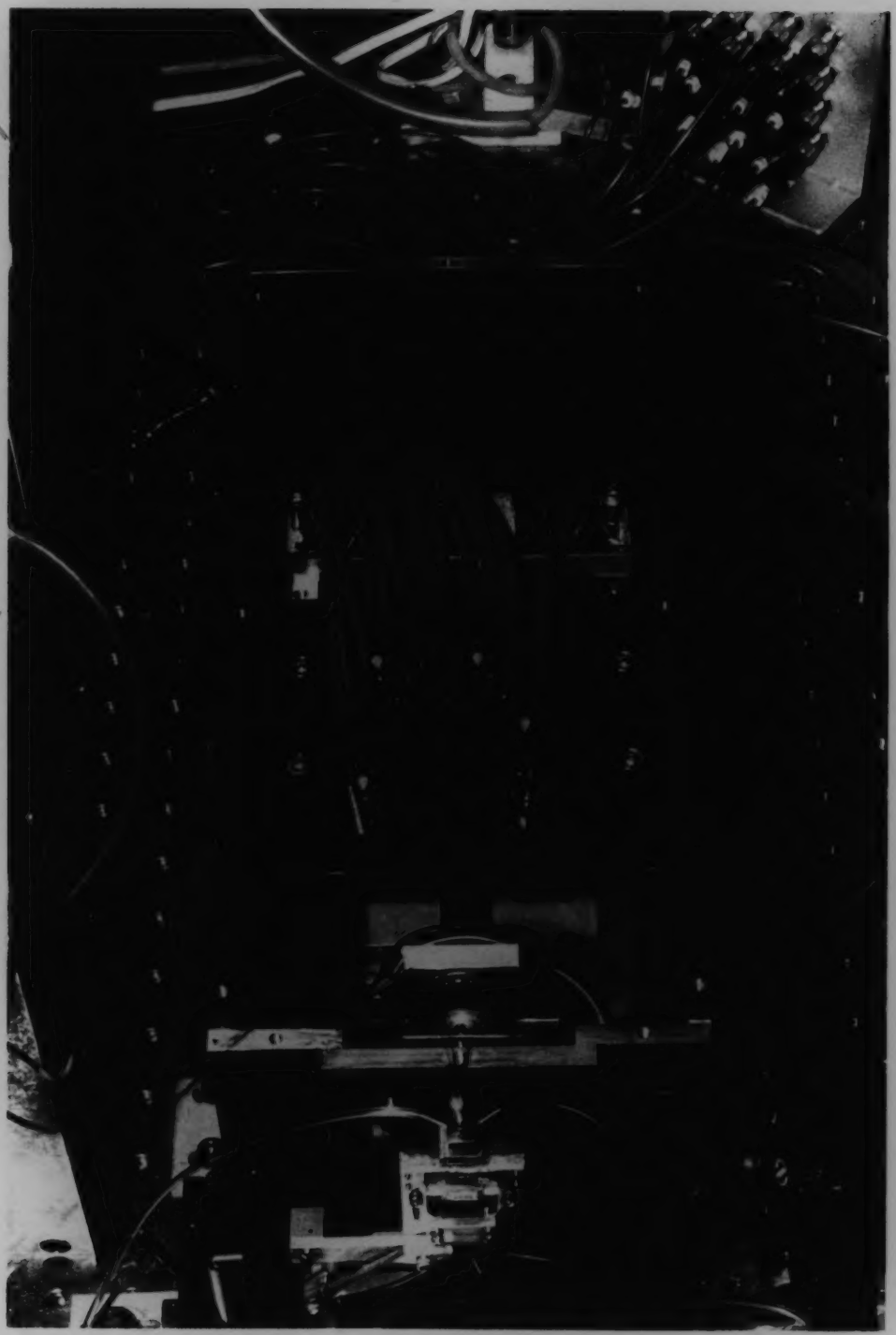
While microscope objectives have high collection efficiencies, they are better suited to image objects that are much smaller than the lens diameter. The 16-capillary cuvette was 4.8 mm wide and its imaging resulted in a distortion of the image plane. Field curvature required a parabolic arrangement of the GRIN lenses shown in Figure 2.18. The marginal rays from the peripheral capillaries (1 and 16) were focused closest to the object; they were 11 cm away from the microscope objective. The on-axis rays from the middle capillaries (8 and 9) were focused furthest away at 17 cm from microscope objective.

2.3.2.2. Avalanche Photodiode Cooling

Regulating the APD temperature at -20.0 ± 0.1°C was important because of the dependence of breakdown voltage on temperature. Small shifts in temperature caused drifts in the baseline levels as the dark current and internal gain varied with temperature.

Frost formation was also another problem when cooling APDs to low temperatures. Ice formation on the APDs leads can create short circuit pathways. To avoid this problem, the leads of the APDs were coated with a dielectric medium (crimson nail polish, Del Laboratories Canada, Inc., Barrie, Ontario). In earlier experiments, the APDs were cooled to -30°C to reduce the dark current levels. Despite the flow of nitrogen through the Styrofoam box, drifts in the APD signal were observed when the box was cooled for long periods. These fluctuations in the baseline were attributed to

Figure 2.18. The Arrangement of the GRIN Lenses.
The orange fiber optic cables terminate at the black GRIN lenses in the center of the photo. The GRIN lenses are mounted in the black holders, which are adjustable along the top of the black platform. At the very bottom of the photo is the cuvette holder with buffer tubing. Just above the cuvette, near the yellow and red wires, is the filter wheel. The GRIN lenses, filter wheel and collection microscope objective (not seen) are housed in a black aluminum box.



moisture leaking into the box. Water in the box would condense onto the U-block and the RTD temperature sensor. Moisture may cause the sensor to ground to the aluminum U-block, affecting the reading and the temperature regulation. Operating the APDs at -20.0°C resulted in slightly higher dark current levels but eliminated the drift in the baseline. The dew point of water is 378 ppm at -30°C and its dew point is 1020 ppm at -20°C [23]; less ice is condensed on the U-block at -20°C.

2.3.2.3. Avalanche Photodiode Electronic Circuitry

The detector apparatus, with the high gain of 10^9 V/A, was very sensitive to noise. Both noise from ground loops and environmental electromagnetic noise dominated the first generation 16-capillary instrument detector.

Ground loops in the electronic circuitry were a major problem. Initially, all of the amplifier grounds, +15 V and -15 V inputs from the external power supply were connected in series. When the ground and power leads of the biasing HV power supplies and amplifiers were connected, this serial connection led to a ground loop. The wires acted as antennas and an audible high frequency sound was heard when amplifiers were turned on. The frequency of the noise was ~1 kHz and lead to each amplifier drawing twice the amount of current as with no load (30 mA per amplifier, normal current draw *vs.* 60 mA per amplifier, with ground loop). To eliminate this ground loop, the amplifiers were connected into 4 daisy-chains of 4 amplifiers as shown in Figure 2.19. The 4 chains were then connected together at a "star" point, the first point where the grounds, +15 V and -15 V inputs connect to the electronic circuitry. A similar ground loop occurred when all the low-pass filters power inputs were connected in series. The low-pass filter grounds, +15 V and -15 V inputs were also daisy-chained in groups of four and the four chains connected to the star point.

Placing the entire setup on a copper sheet (60 cm × 30 cm), which served as a ground plane, also reduced the occurrence of ground loops. Ground cables were switched from relatively thin 18-gauge wire to fatter braided cable to reduce their internal resistance and induction. The bias voltage coaxial cables were grounded at the HV end to prevent

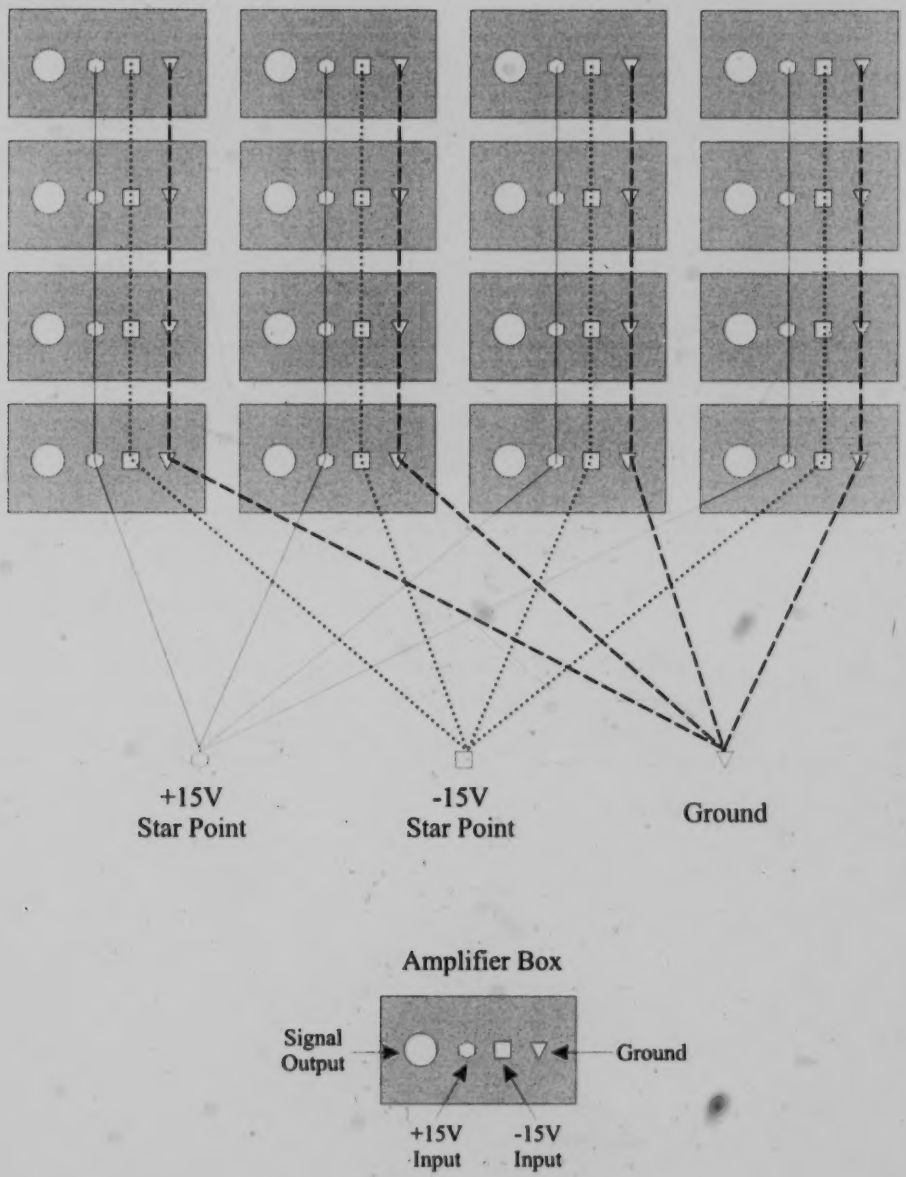


Figure 2.19. Input Connections for Transimpedance Amplifiers
The amplifiers were connected as four groups of four. The inputs for +15 V, -15 V, and the grounds are connected in groups of four, also called a daisy-chain of four. The daisy-chains met at a common star point.

ground loops if the APD end ground shifted. The coaxial cable leads were kept as short as possible.

Environmental 60 Hz noise was observed with a peak-to-peak amplitude of 2 V when the APDs were biased. Shields within shields, as shown by the dashed lines in Figure 2.13, were constructed to isolate the HV APD power supplies, the amplifiers, the low-pass filters and APDs from the environmental interference. As the first shield, the biasing HV power supplies were placed behind a steel barrier to reduce coupling with rest of electronic circuitry. The transimpedance amplifiers were pre-packaged in aluminum boxes, which served as another layer of shielding. As a third shield, the APD leads from the U-block were encased in a cylindrical piece of Ad-Mu foil (Ad-Vance Magnetics Inc., Rochester, IN) as shown in Figure 2.12. The Ad-Mu foil, U-block and copper heat sink were grounded. The exterior of the Styrofoam APD box was shielded by aluminum foil. The entire electronic setup was encased in a grounded stainless steel box (60 cm × 30 cm × 30 cm). All this shielding reduced the peak-to-peak amplitude of the 60 Hz noise from 2 V to 20 mV.

High frequency ripple on the ±15 V inputs for the amplifiers, the biasing voltage, and the low-pass filters were AC decoupled using passive low-pass filters with a time constant of 0.1 ms [12]. We also observed a high frequency noise of ~80 MHz on the signal from our amplifiers. Its peak-to-peak amplitude was small, approximately 3 mV, as the low-pass filters attenuated its amplitude. This high frequency noise may be environmental because it was detected with the instrument off. To reduce this noise, the gain of the low-pass filters was increased from 1 to 10.

The overall background noise of the system, with APDs biased at the values listed in Table 2.4, had a peak-to-peak amplitude of a few mV after all the above grounding and shielding precautions were made. This noise value was three orders of magnitude lower than the noise from the original design of the electronics.

2.3.2.4. Avalanche Photodiode Biasing Voltage

At a constant LED light intensity, the APD signal increased with increasing APD bias. However, the noise in the signal also increased with increasing APD bias. As a

result, the signal-to-noise ratio varied with APD bias. The APD bias with the maximum signal-to-noise ratio was selected as the APD biasing voltage.

The chosen biases are listed in Table 2.4. At the chosen biases, the average APD signal was about 10 mV and the noise in the signal a few mV.

Table 2.4. APD BIAS at -20°C

APD	Breakdown Voltage (V)	Biasing Voltage (V)	Maximum S/N
1	-157	-154.5	8
2	-220	-211.4	7
3	-218	-208.0	8
4	-217	-209.2	5
5	-214	-207.5	6
6	-215	-213.4	11
7	-213	-211.3	9
8	-221	-212.4	8
9	-218	-210.0	6
10	-220	-210.4	6
11	-223	-215.1	9
12	-209	-199.4	7
13	-223	-215.4	6
14	-217	-207.0	8
15	-216	-207.8	8
16	-213	-203.1	6

2.3.2.5. Chopper/Filter Wheel and Data Acquisition

Initial sequencing runs on the 16-capillary instrument were performed using a first generation locally constructed chopper/filter wheel. These chopper/filter wheels did not have a feature that synchronized the wheels if they fell out of alignment. The filter wheel was much heavier than the chopper wheel, and sometimes its stepper motor would miss a step. After missing several steps, the wheels would fall out of synchronization. The data would be acquired with the wrong bandpass filter and laser combination or with the filter holder blocking the fluorescence. The present chopper/filter wheel has an auto-align function. If one wheel fell out of synchronization, the leading wheel waited one rotation to synchronize its rotation with the lagging wheel. A more detailed description of the current data acquisition program can be found in D. Lewis' thesis [19]. This chopper/filter wheel was used for the wt M13 sequencing runs.

2.3.3. DNA Sequencing

2.3.3.1. Single-Stranded Template Sequencing

Figure 2.20 shows a portion of a ssM13mp18 sequencing sample run simultaneously on 16 capillaries. The section was 309-325 bases after the sequencing primer with the sequence, "AGGGTGGTTTTTCTTTT." The peaks were blue-C, green-A, black-G, and red-T. The poor resolution of the quintet and a quartet of T-terminated fragments was due to the gel chemistry and not the 16-capillary instrument. The T-terminated peaks were the most intense signals, twice as high as all the other termination peaks except in capillaries 11, 15 and 16, where the G-terminated peaks' were almost as strong as the T-terminated. If it was due to the green laser alignment, then capillaries 12 through 14 would also see similar G-terminated peaks' signal increase. This feature may arise from an imperfection in the 580DF10 (TAMRA) bandpass filter. The filter aligned the fluorescence images from capillaries 11, 15 and 16 better to detection optics.

2.3.3.2. Cloned Wild-Type M13 Sequencing

In the wild-type M13 sequencing exercise, the data from about 25% of the electropherograms over 40 runs (1 run = 16 electropherograms), generated resolvable sequencing peaks. Of the successful electropherograms, only about one quarter (6% of all the electropherograms) had adequate signal levels to be analyzed by the base-calling software BASS (see Section 5.2.7).

Figure 2.21 shows a successful run from wt M13. This electropherogram was from capillary 10 from file: sb960528-1 and data processed with the base-calling program. The sample was double-stranded plasmid template clone 45, run at 35°C. This clone was a fragment of wtM13 starting at 4119 bases. Sequence from over 500 bases was generated between 40 to 220 minutes. The signal-to-noise ratio was low, typical of most of the electropherograms from the first generation instrument. This low signal-to-noise ratio was not acceptable for DNA sequencing. The BASS program had difficulty

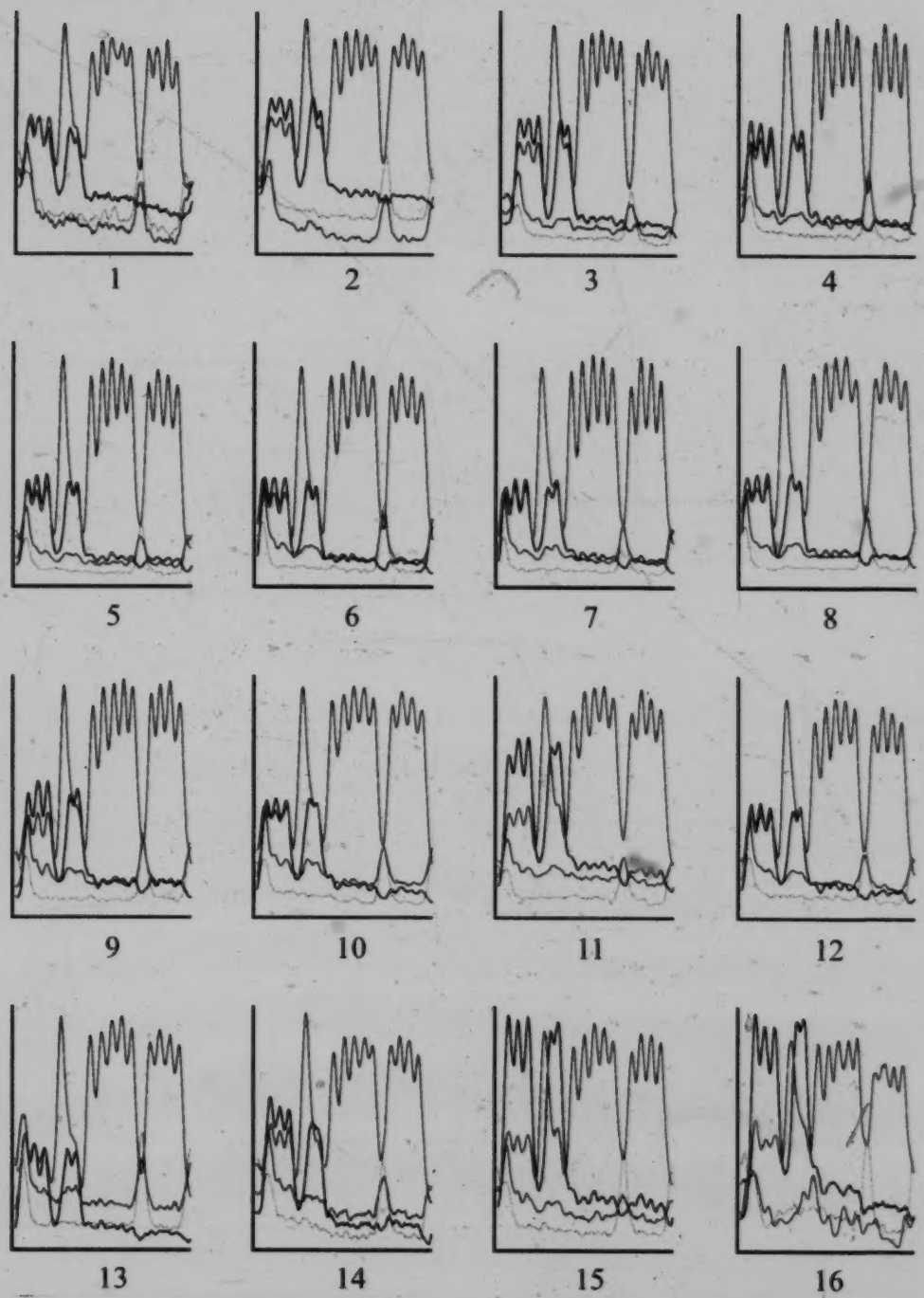


Figure 2.20. A 16-Capillary Run of Single-Stranded M13mp18.
A portion of 16 capillaries simultaneously run, 309-325 bases after the primer. Run at 200 V/cm, room temperature, on 4%T gel. Green=A, blue=C, black=G, red=T.

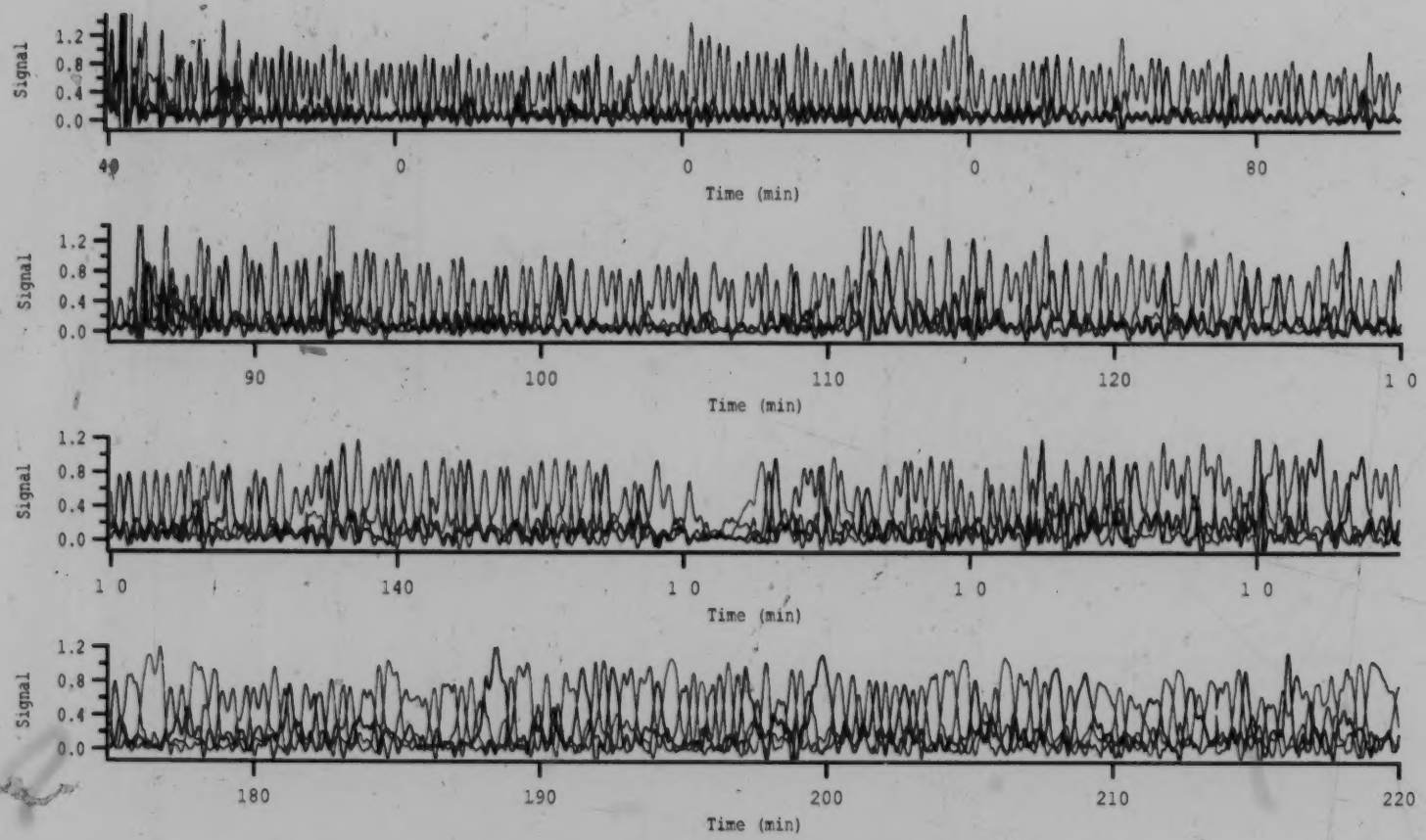


Figure 2.21. Wild-Type M13 Sequencing Run.
Sequence from over 500 bases of cloned wt M13. Run on 6%T polyacrylamide gel at 150 V/cm and 35°C.

base-calling the electropherograms with low signal levels. In order to obtain long sequencing read lengths with high sequencing accuracy, the signal-to-noise level needed to be increased.

In sequencing wt M13, about 25% of the all electropherograms were completely ruined by sample artifacts that I called lumps and stretches. A small artifact can be seen in Figure 2.21 at 150 min. In this case, the gap was very minor because only 3 bases of sequence were lost and it did not affect the ability to read most of the sequence. This artifact, a stretch, was attributed to secondary DNA structure. The sequencing artifacts are addressed in more detail in Chapter 7.

Preliminary sequencing results with 4% T polyacrylamide gel and capillary heating resulted in a large percentage of capillary failure. As a result, the higher gel concentration of 6%T was used for the heated wt M13 runs with a slightly improved success rate. This lab has pioneered capillary DNA sequencing at elevated temperatures [24]. Heating the capillaries is necessary during the sequencing run because it melts the intra-strand secondary structure in sequencing fragments that 7 M urea could not remove. Heating is also important because sequencing read length increases with increasing temperature [24, 25].

However, out of the 40 runs on the 16-capillary instrument, about 50% of the 6%T gel electropherograms failed due to bubbles in gels, capillary failure after sample injection and poor peak resolution. A drop in the capillary's current and the primer peaks appearing very late in the run, indicated capillary breakdown. The poor resolution was attributed to poor coating of the gel to the capillary wall, resulting in residual electroosmotic flow in the capillary that opposes the electrophoretic mobility of the DNA. The current drop and the poor resolution were the possible result of overloading the capillary column with sample.

These types of gel failures were also seen when capillaries were run at room temperature but the percentage of failures increased with capillary heating. Chapter 3 presents an improved method for heating the capillaries.

The gels were polymerized *in situ* and the capillaries were replaced for each run. Despite preparing the gels similarly each time, there was an obvious variation from batch

to batch. Refilling the capillaries with the same gel batch should reduce variations in gel quality. Refilling multiple capillaries simultaneously is addressed in Chapter 3.

2.4. CONCLUSIONS

I have presented the preliminary results from the 16-capillary DNA sequencer based on a micromachined sheath flow cuvette and APD detectors. Advantages of this glass micromachined cuvette are its low cost, rapid large-scale production, low light scatter, and the ability to spatially fix the capillaries. Quartz cuvettes, like those used in the single capillary and the tapered 5-capillary cuvettes, are relatively expensive to purchase and can take months to manufacture commercially. These glass cuvettes were less expensive because they are batch-fabricated. New cuvettes can be prepared in only a few days.

Four-color DNA sequencing on all 16 capillaries was successfully demonstrated. However, the instrument suffered from low signal-to-noise levels. Increasing the amount of sample loaded onto the capillary was not acceptable. While it may improve signal-to-noise levels, it also deteriorates the resolution and increases the costs of the sequencing reaction.

Sequencing the cloned DNA revealed the need to study further the many parameters of capillary DNA sequencing such as the effect of heating the capillaries, gel polymerization inconsistencies and sequencing artifacts. Combining these problems with poor signal-to-noise of the sequencer resulted in a low fraction of useful electropherograms generated by the wt M13 exercise.

Some of the problems revealed in this chapter will be addressed in Chapter 3, where I describe an improved version of the 16-capillary sequencer. This chapter addresses problems such as the poor signal-to-noise ratio of the APD detectors, capillary heating and simultaneous sample injection and capillary gel refilling. Chapter 4 details the methodology used to improve the sensitivity of the APD detectors. Chapter 5 is an evaluation of the sequencing capability of the second generation 16-capillary instrument and Chapter 6 examines the base-calling software accuracy. Chapter 7 describes some of the sequencing artifacts observed and their elimination.

2.5. REFERENCES

(1) Ghandhi, S. K. *VLSI Fabrication Principles*; John Wiley and Sons, **1983**.

(2) Kovacs, G. T. A.; Petersen, K.; Albin, M. *Analytical Chemstry* **1997,** *68*, 407A-412A.

(3) Kopp, M. U.; Crabtree, H. J.; Manz, A. *Current Opinion in Chemical Biology* **1997,** *1*, 410-419.

(4) Manz, A.; Widmer, H. M. *Sensors Actuators B* **1990,** *1*, 244-248.

(5) Harrison, D. J.; Manz, A.; Fan, Z.; Ludi, H.; Widmer, H. M. *Analytical Chemistry* **1992,** *64*, 1926-1932.

(6) Jacobson, S. C.; Ramsey, J. M. *Analytical Chemistry* **1996,** *68*, 720-723.

(7) Crabtree, H. J. Ph.D., University of Alberta, Edmonton, **1997**.

(8) Optics Guide 5, Melles Griot, **1990**.

(9) Hsing, T. R. In *Fiber Optics*; Daly, J. C., Ed.; CRC Press, Inc.: Boca Raton, Florida, **1984**, pp 195-236.

(10) SELFOC Product Guide, NSG America, Inc.

(11) Dautet, H.; Deschamps, P.; Dion, B.; MacGregor, A. D.; MacSween, D.; McIntyre, R. J.; Trottier, C.; Webb, P. P. *Applied Optics* **1993,** *32*, 3894-3900.

(12) Malmstadt, H. W.; Enke, C. G.; Crouch, A. R. *Electronics and Instrumentation for Scientists*; The Benjamin/Cummings Publishing Company, Inc.: Don Mills, Ontario, **1981**.

(13) Avalanche Photodiodes: A User's Guide, EG&G Optoelectronics, Canada.

(14) EG&G, Photodiode Noise Characteristics, **1997**.

(15) Silicon Avalanche Photodiodes C30902E, C30902S, C30921E, C30921S Data Sheet, RCA ElectroOptics.

(16) Yanisch-Perron, C.; Vieira, J.; Messing, J. *Gene* **1985,** *33*, 103-119.

(17) Ingle, J., J.D.; Crouch, S. R. *Spectrochemical Analysis*; Prentice-Hall, Inc.: Toronto, **1988**.

(18) Swerdlow, H.; Zhang, J. Z.; Chen, D. Y.; Harke, H. R.; Grey, R.; Wu, S.; Dovichi, N. J.; Fuller, C. *Analytical Chemistry* **1991,** *63*, 2835-2841.

(19) Lewis, D. F., Ph.D. Thesis, in preparation, University of Alberta.

(20) Hjerten, S. *Journal of Chromatography* **1985,** *347*, 181-198.

(21) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbour Laboratory Press: Plainview, N.Y., **1989**.

(22) Dovichi, N. J., Lasers In Chemical Analysis, **1994**.

(23) Dew Point Data Sheet, Liquid Carbonic Canada, Ltd., **1980**.

(24) Zhang, J. Z.; Fang, Y.; Hou, J. Y.; Ren, H. J.; Jiang, R.; Roos, P.; Dovichi, N. J. *Analytical Chemistry* **1995,** *67*, 4589-4593.

(25) Carrilho, E.; Ruiz-Martinez, M. C.; Berka, J.; Smirnov, I.; Goetzinger, W.; Miller, A. W.; Brady, D.; Karger, B. L. *Analytical Chemistry* **1996,** *68*, 3305-3313.

Chapter 3: 16-Capillary Sequencer, Version 2.0

3.1. INTRODUCTION

Capillary DNA sequencers are easier to automate than the commercially available slab gel sequencers. Slab gels require manual pipetting of samples into the gel well and new gel plates must be poured between runs. In contrast, samples can be injected automatically and the gel matrix pumped through the capillaries between runs. PE Applied Biosystems sells a single-capillary DNA sequencer that is fully automated [1]. An automated multiple-capillary instrument would increase sample throughput tremendously.

A second-generation version of the 16-capillary DNA sequencer, described in this chapter, treads a few steps closer to an automated multiple capillary instrument. The changes to the instrumentation are described in two sections: the separation apparatus and the detection apparatus. In the separation apparatus, the sample injection was simplified and the capillary heater was modified to improve the quality of data. A 16-capillary gel refiller was added to the instrumentation. In the detection apparatus, the avalanche photodiode (APD) electronic circuitry was altered to increase sensitivity and reduce the noise of the signal. The modifications to the instrument improved its sensitivity and facilitated its use.

3.1.1. Capillary Gel Refiller

The flow of liquids in pipes can be described by Poiseuille's equation:

$$\text{Volume Flow Rate} = \frac{\Delta V}{\Delta t} = \frac{\pi \cdot \Delta P \cdot d^4}{128 \cdot L \cdot \eta} \qquad (3.1)$$

where, $V = \pi d^2 L/4$ is the pipe volume, t is the time, ΔP is the pressure drop across the pipe, d is the diameter of the pipe, L is the pipe length and η is the viscosity of the fluid. Rearranging the Equation 3.1 and substituting the formula for the volume of a cylinder gives the time needed to fill a capillary of length L and inner diameter d:

$$\Delta t = \frac{32 \cdot L^2 \cdot \eta}{d^2 \cdot \Delta P} \qquad (3.2)$$

From Equation 3.2, the time required to refill a capillary is reduced by decreasing the capillary length, decreasing viscosity of the gel, increasing the pressure difference, or increasing the inner diameter of the capillary.

To achieve short refill times, low viscosity gels that can separate DNA sequencing fragments are highly desirable. A gel that exhibits these properties is a poly-N,N-dimethylacrylamide (DMA) gel. PE Applied Biosystems has developed a DMA polymer [2] that has viscosity of ~500 cpoise. This polymer is also interesting because it does not need to be covalently bound to the capillary walls.

The times to refill a capillary, 50 μm ID, 60 cm long at 250 psi (1700 kPa), with various fluids are listed in Table 3.1.

Table 3.1. Refill Times of Various Fluids

Fluid	Viscosity at 20°C (cpoise)	Refill Time
water*	1	3 s
ethylene glycol*	20	1 min
olive oil*	84	4 min
ABI-DMA†	500	23 min
glycerin*	1500	1 h

* [3] and † [4]

There are a number of ways to apply pressure to pump the sequencing gels through the capillaries. We chose to apply gas pressure to the fluid. Helium gas was used to pressurize the 16-capillary gel refiller because it has a lower solubility parameter with increasing pressure compared to neon and nitrogen [5]. Helium has the lowest molar ratio, x_A, moles of gas per mole of liquid. In water, x_A is equal to: 0.000012 (N_2), 0.0000084 (Ne) and 0.000007 (He) at 25°C and 1 atm (14.7 psi) [6].

3.2. EXPERIMENTAL

3.2.1. Separation Apparatus

3.2.1.1. Capillary Gel Refiller

A miniature pressurized gel refilling device ("the bomb") was made by John Bercovitz at Lawrence Berkeley National Laboratories (Berkeley, CA) [7]. Figure 3.1 shows a drawing of the device. The bomb was made from 3 stainless steel parts: the

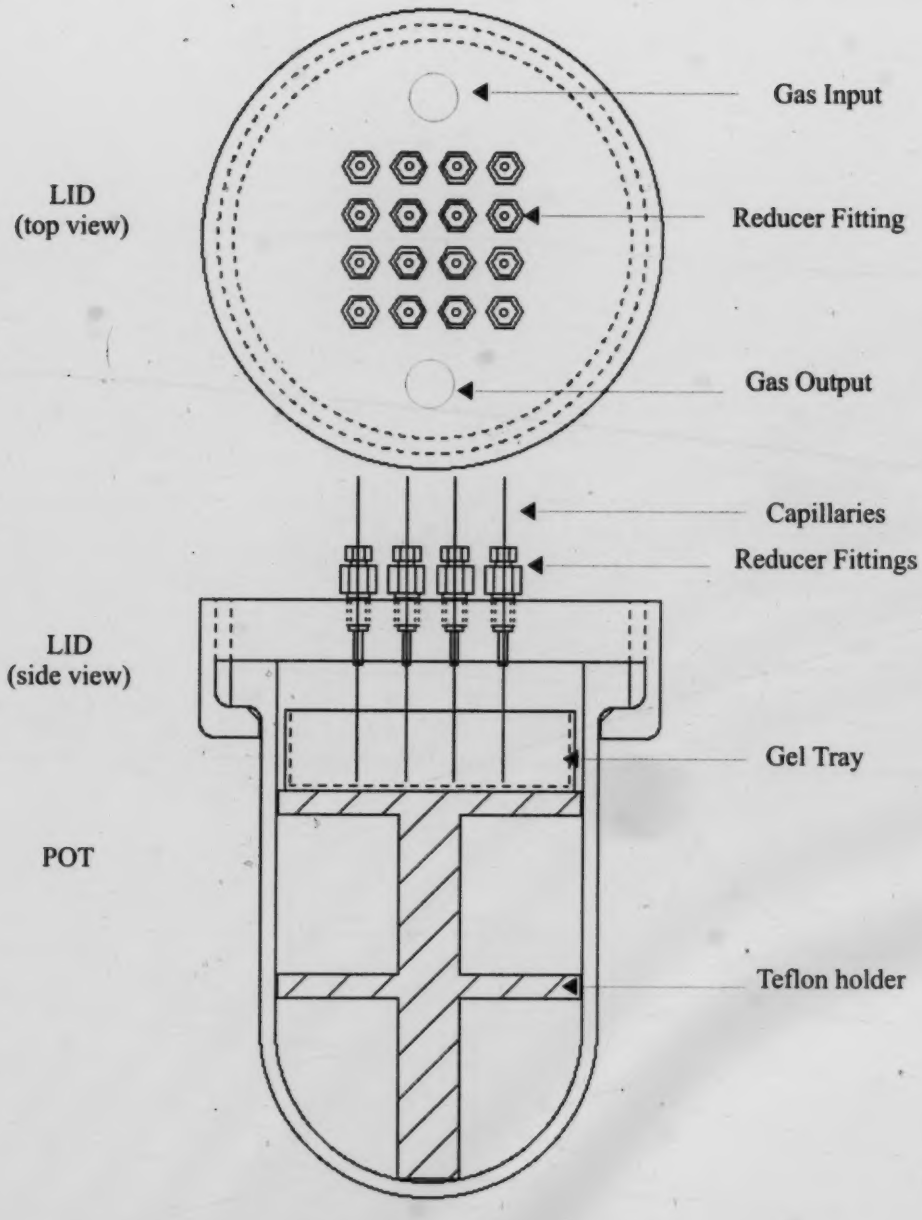
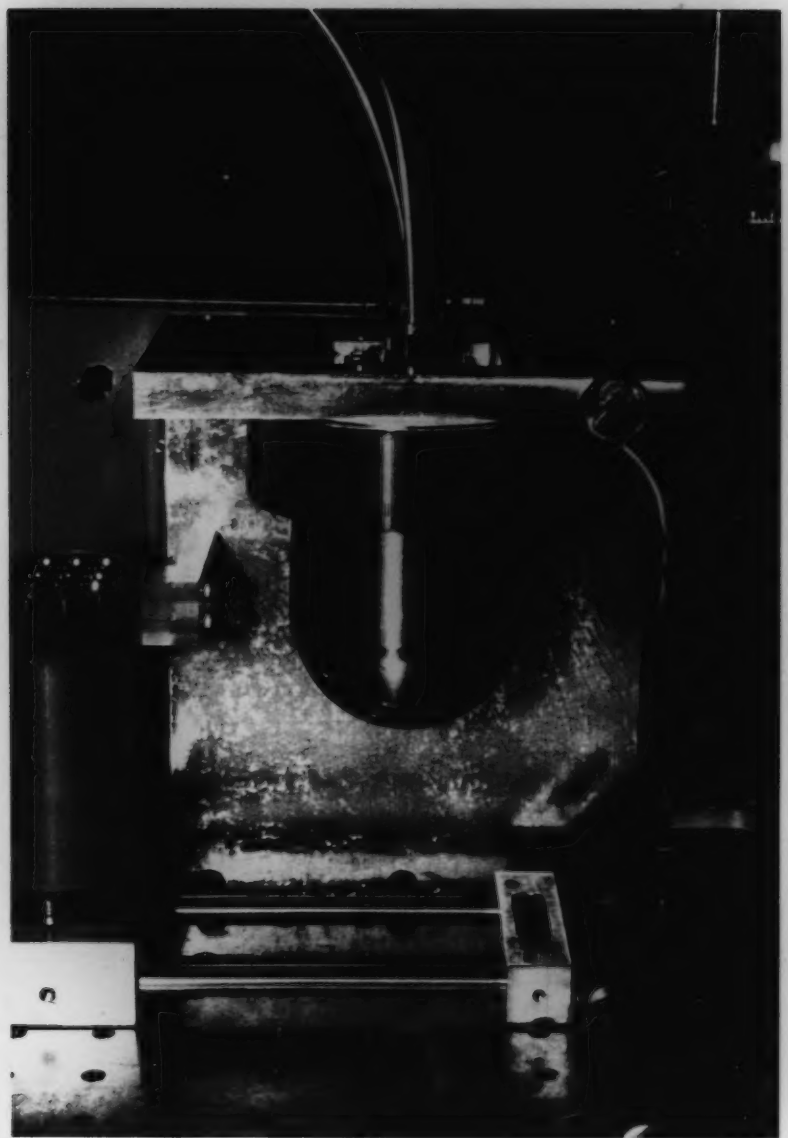


Figure 3.1. The Capillary Gel Refiller.
Capillaries are inserted through reducer fittings mounted in lid. The capillary ends are submerged in the gel tray, positioned by the Teflon holder. The refiller is pressurized by helium gas; the input is shown in the top view of the lid.

Figure 3.2. Top View of the 16-Capillary Gel Refiller.
Sixteen capillaries thread through the stainless steel reducer fittings mounted in a 4x4 array of the lid of the gel refiller. The input and output gas lines are visible from the top. The gel refiller lid is mounted into an aluminum bracket.

Figure 3.3. The 16-Capillary Gel Refiller.
The 16-capillary gel refiller is mounted onto an aluminum bracket. The bottom stainless steel pot threads onto the lid with the ring nut. The steel tubing at the top are the input and output gas lines. The sample injector carousel with gold-plated microtitre plate can be seen on the left of the photo.



bottom pot, the lid and, a ring nut that joins the pot to the lid. Inside the bomb pot was a Teflon holder that positions the gel tray at the specific height for the capillary ends.

Sixteen capillaries were fed into a 4×4 array of reducer fittings (IZR1.5 I-reducer, 1/6" to 1/32", Chromatographic Specialties Ltd., Brockville, Ontario) with Vespel ferrules (FS.4 Adaptor 1/32", Chromatographic Specialties) screwed into the lid of the stainless steel bomb gel pressure refilling apparatus. Figure 3.2 shows a top view of the capillary arrangement. During electrophoresis, the bomb pot was removed and the bomb lid served to hold the capillaries for sample injection. The bomb lid had input and output valves to connect gas pipe lines. The output pipe line had a muffler for exhaust, a safety port with pressure release valve and a pressure gauge.

Between the sequencing runs, the capillary ends were placed in the gel tray and the bomb pot was attached to the bomb lid. The bomb was operated at 250 psi but could be operated as high as 500 psi. The bomb was tested to 1000 psi for 30 min [7]. A photo of the apparatus refilling the capillaries is shown in Figure 3.3.

3.2.1.2. Sample Injector

The sample injector consisted of a Plexiglas insulator, the HV power supply, a gold-plated 18-well microtitre plate and a sample carousel. Figure 3.4 shows a schematic of the sample injector and Figure 3.5 shows a photo.

The capillaries were threaded through a Plexiglas insulator fixed to the underside of the bomb lid. The Plexiglas insulator and nail polish (Del Laboratories Canada, Inc., Barrie, Ontario) on the bottom of the bomb lid prevented arcing of the voltage to the steel bomb lid. The HV cable connected to a banana plug recessed inside the Plexiglas insulator. The HV cable was disconnected for gel refilling. A platinum electrode was soldered to the other end of the banana plug. The electrode was positioned to fall into an extra well of the gold-plated microtitre plate.

A 96-well polycarbonate microtitre plate (AS-5260, Applied Scientific, San Francisco, CA) was plated with gold in an evaporation chamber (model SE 600 RAP, CHA Industries, Palo Alto, CA) or in an Ar^+ plasma sputtering system (model 4410, Perkin-Elmer, Mississauga, Ontario). Figure 3.6 shows a photo of the gold-plated 96-

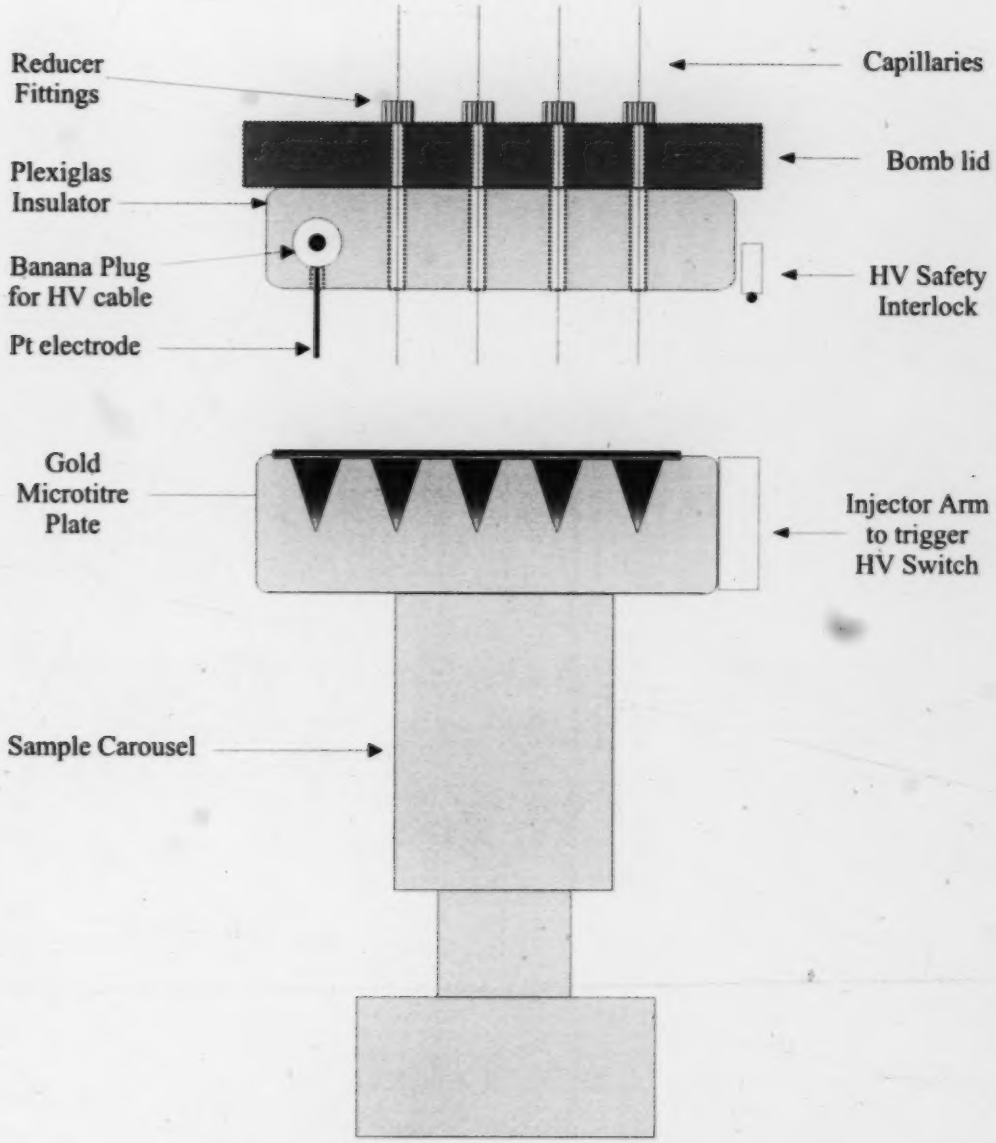
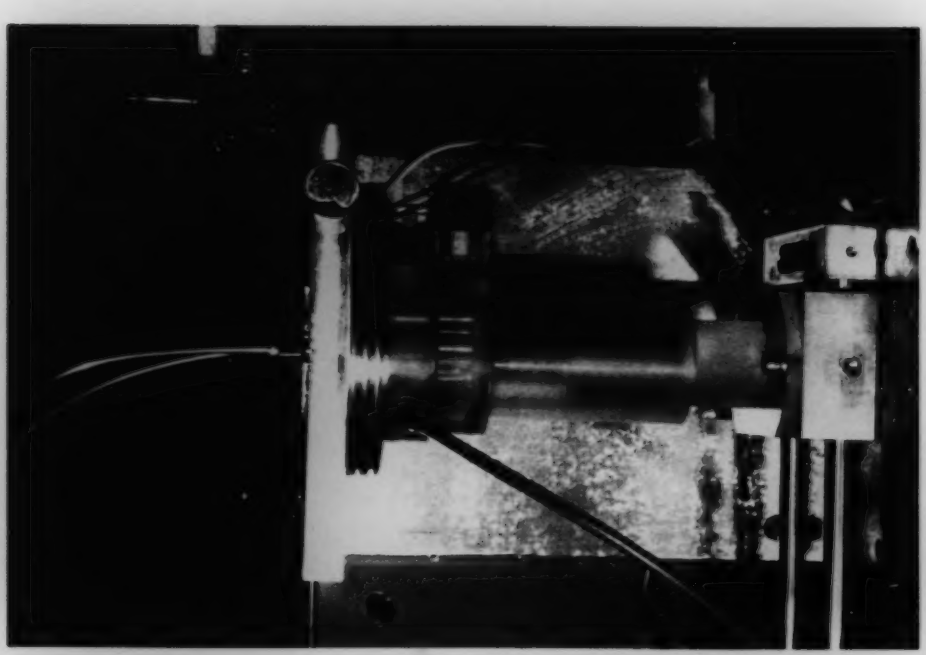


Figure 3.4. The Sample Injection Apparatus.

The capillaries are inserted through the reducer fittings in the stainless steel bomb lid and the holes in the Plexiglas insulator. The banana plug in the Plexiglas insulator connects the HV cable to the Pt electrode. The electrode is positioned to fall into an extra well in the gold microtitre plate. The sample carousel is spring-loaded, moving up to meet the capillaries. An arm on the side of the carousel triggers the HV switch.

Figure 3.5. The Sample Injector

The capillaries thread through the top of the stainless steel bomb lid. The Plexiglas insulator, 0.5" thick, mounts on the underside of the bomb lid. The pink high voltage cable connects to a platinum electrode in the Plexiglas insulator. The sample carousel houses the gold-plated microtitre plate. The carousel is spring-loaded and the arm on the right of the carousel triggers the safety interlock (the yellow and green wires) on the side of the aluminum mount, shown in the photo on the right.



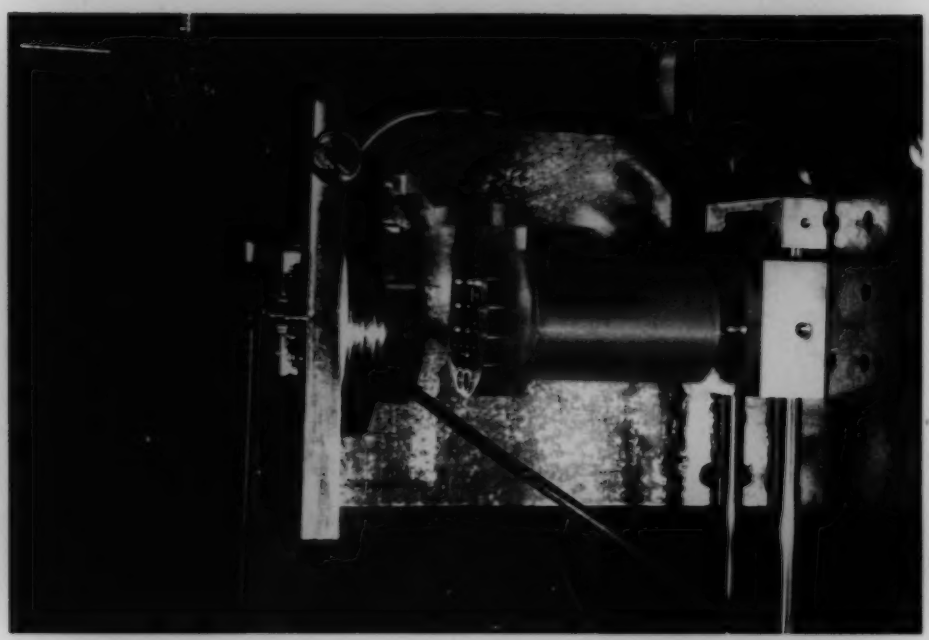
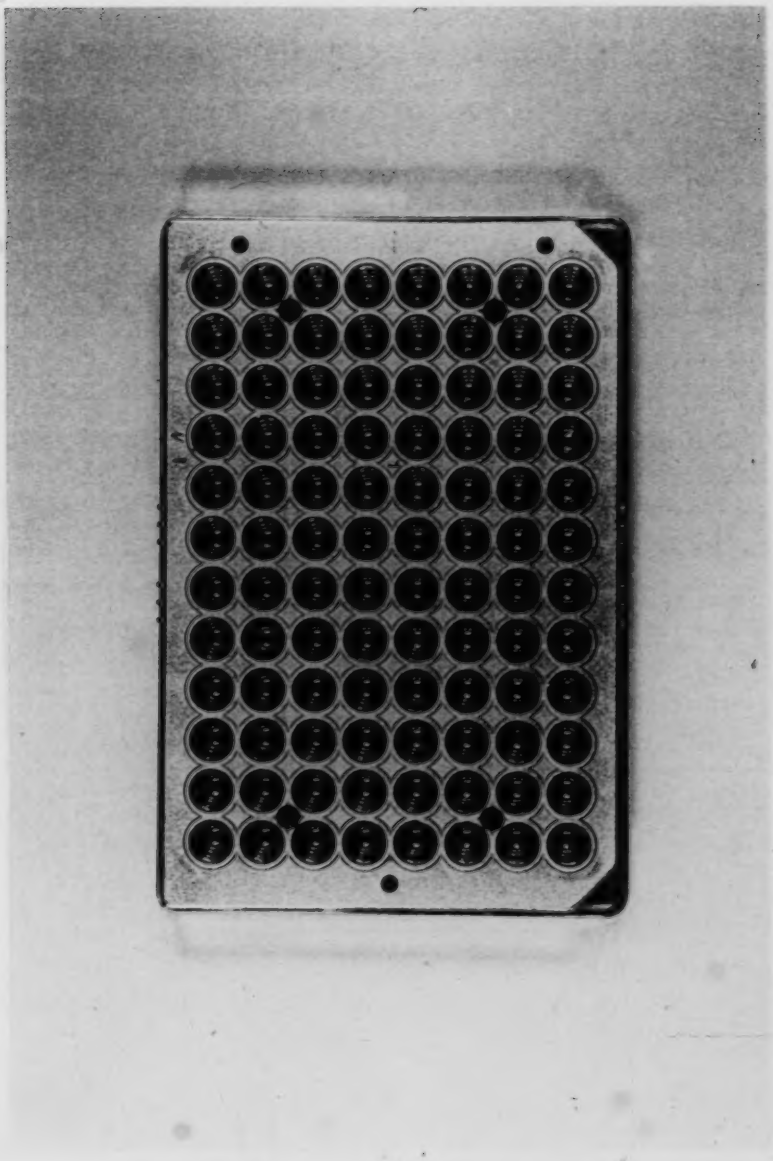


Figure 3.6. The 96-Well Gold-Plated Microtitre Plate.
Gold, 300 nm thick, plated onto a polycarbonate microtitre plate.



well microtitre plate. For the 16-capillary instrument, the microtitre plate was cut into a 4×4 array plus an extra well for the Pt electrode.

The gold microtitre plate was housed in a spring-loaded Plexiglas sample carousel. The carousel moved up and down to meet the capillary ends and the Plexiglas insulator. During injection, the gold-plated microtitre plate was sandwiched between the sample carousel and the Plexiglas insulator, isolating the microtitre plate from the user. When the carousel mated with the Plexiglas insulator, an arm on the carousel triggered a HV safety interlock. The HV switched off when the carousel was lowered.

3.2.1.3. Peltier Capillary Heater

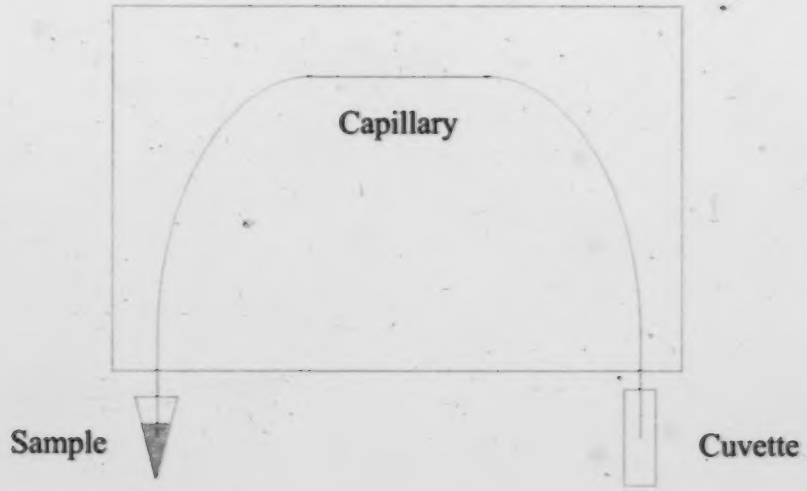
A diagram of the capillary heater is shown in Figure 3.7 and a photo in Figure 3.8. Eight Peltier devices (CP 1.4-7-10L, Melcor, Materials Electronics Products Corp., Trenton, NJ) were serially connected and sandwiched between two copper plates (28 cm × 14 cm × 1 mm). One face of a copper plate was covered with a 0.5 mm thick sheet of Teflon to prevent conduction through the metal if a capillary broke during electrophoresis. The capillaries were sandwiched between a piece of foam (28 cm × 14 cm × 1 cm) and the Teflon-covered copper face. The entire assembly was housed in a Plexiglas box.

The Peltier devices were heated by applying constant DC voltage from a power supply (MSK 10-10M, Kepco, Flushing, NY). The temperature was monitored with a thermocouple (Digi-Sense Thermometer, Cole-Parmer Instrument Company, Vernon Hills, IL). About 10 cm from the injector end and 5 cm from the cuvette end of the capillaries were not heated.

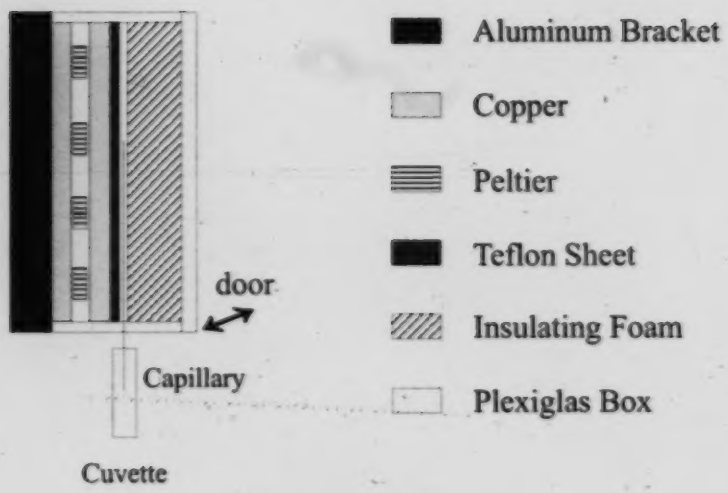
3.2.2. Detector Apparatus

3.2.2.1. Optics

The original filter set consisted of 540DF10, 560DF10, 580DF10 and 610DF10 (Omega Optical, Brattleboro, VT) for the dyes FAM, JOE, TAMRA and ROX respectively. In this version of the instrument, the FAM and JOE filters were replaced with bandpass filters 515DF20 and 540DF40 (Omega Optical Inc.).



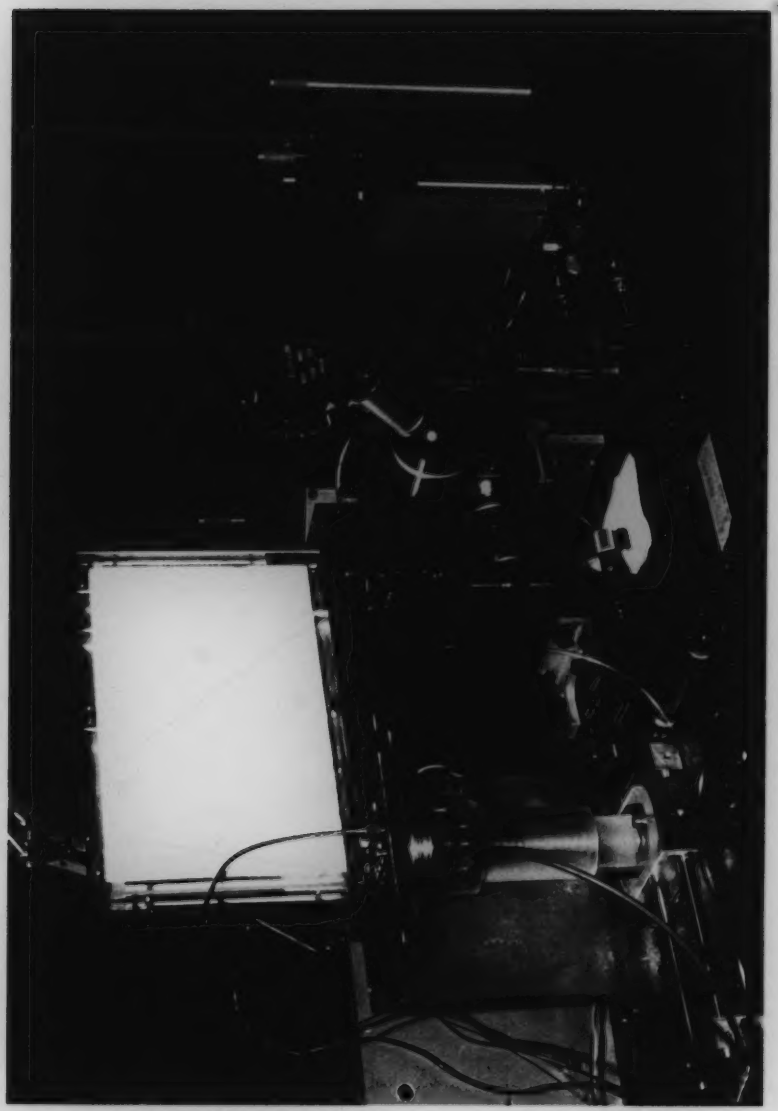
a) Front view



b) Side view

Figure 3.7. Peltier Capillary Heater.

Figure 3.8. The Peltier Capillary Heater.
The white rectangle is the foam face of the Peltier capillary heater encased in a Plexiglas case. The heater straddles the sample injector on the left of the photo and the black optics box in the center of the photo. The capillaries exit the bottom right of the capillary heater into the cuvette.



3.2.2.2. Avalanche Photodiode Cooling

A new box was constructed out of copper circuit-board, shown in Figure 3.9, to thermally insulate and electromagnetically shield the APDs [8]. A second box, also made from the circuit-board, was attached to the side of the box with APD leads. The APD leads were directly soldered to the face of the second box to reduce microphonic noise generated from vibrations.

Inside the APD box, the aluminum block holding the APDs was cut from a "U" to an "L" shape to reduce its mass and reduce the energy required to cool the block. The Peltier device (CP-1.4-71-10L, Melcor, Materials Electronics Products Corp.) was cooled using a PID (Proportional/Integral/Derivative) temperature controller (model SE5020, Marlow Industries, Dallas, TX) and a T-type thermocouple temperature sensor. The temperature controller was used with settings: 4 A maximum current, -20.0°C set point, 8.0°C proportional band, 30 s integration time, and no derivative action. The Peltier device was placed on top of a water-cooled copper-heat sink.

3.2.2.3. Avalanche Photodiode Electronic Circuitry

As in Chapter 2, specially modified transimpedance amplifiers (model 341, Analog Modules, Inc., Longwood, FL) conditioned the signal from each APD (C30902S, EG&G Canada, Vaudreuil, Quebec). The gain in the amplifiers was increased from 10^9 V/A to 10^{10} V/A by changing the resistor in the amplifier feedback loop from 1 GΩ to 10 GΩ (Figure 3.10). The signal was low-pass filtered with gain of 20 and 8 Hz cutoff.

The entire electronic circuit was rewired to reduce ground loops. The coaxial cable lengths were minimized and all the components were carefully grounded and shielded. The ground, +15 V, and -15 V input connections for the transimpedance amplifiers and low-pass filters were daisy-chained into groups of 2 and the chains connected at a star point. The amplifiers, the APD HV power supplies, and low-pass filters were placed on a copper ground plane inside an aluminum box. This aluminum box and the APD box were placed on a copper ground plane inside a steel box with the dimensions, 35 cm × 35 cm × 15 cm. Figure 3.10 shows the electromagnetic shields for the circuitry.

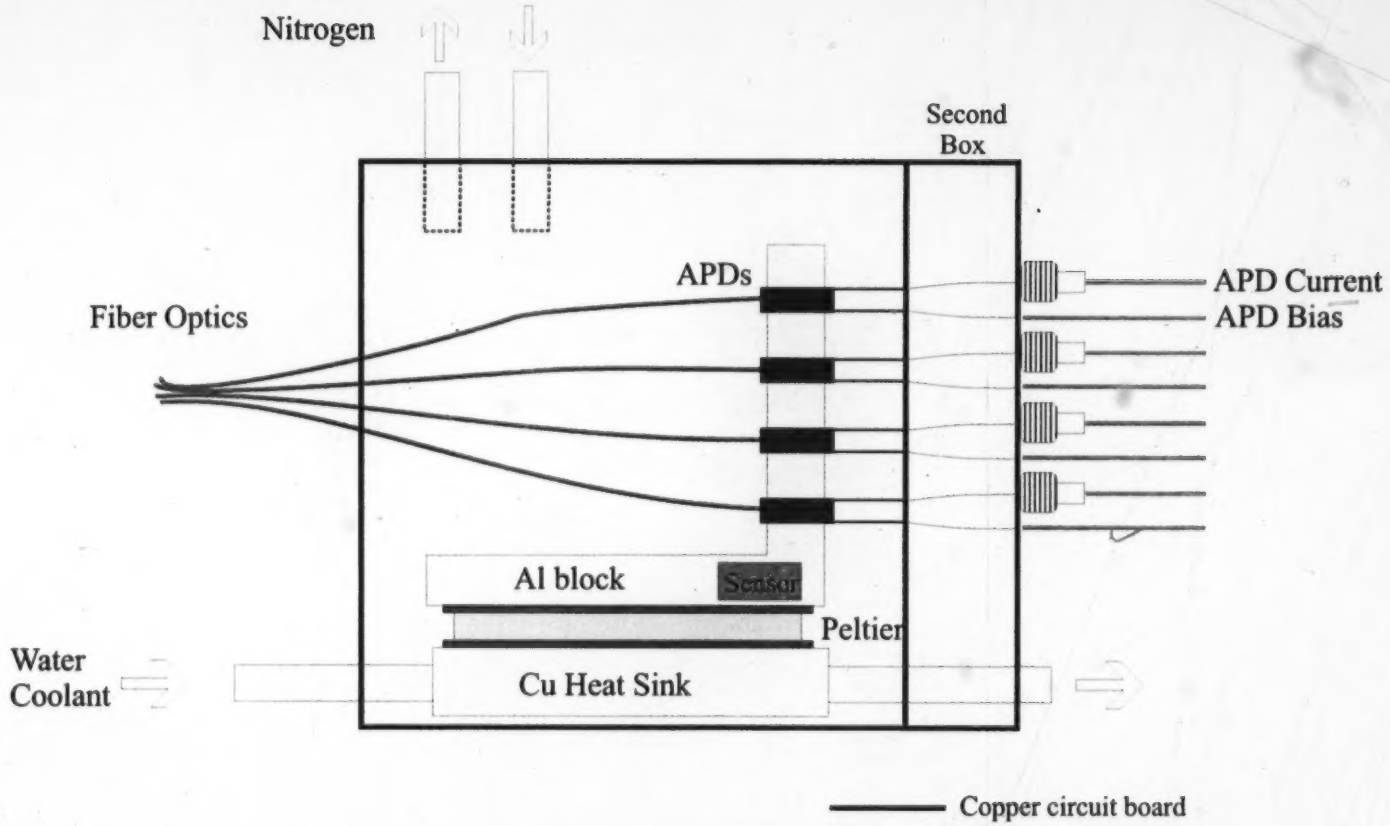


Figure 3.9. APD Cooling Box.

The APD box is constructed from copper circuit board. As before, the APD are mounted in an aluminum block, and cooled to -20°C by the Peltier. The APD leads are soldered directly to the right face, which connects to a second box. The second box, made from copper circuit board, serves to electromagnetically shield the APD leads.

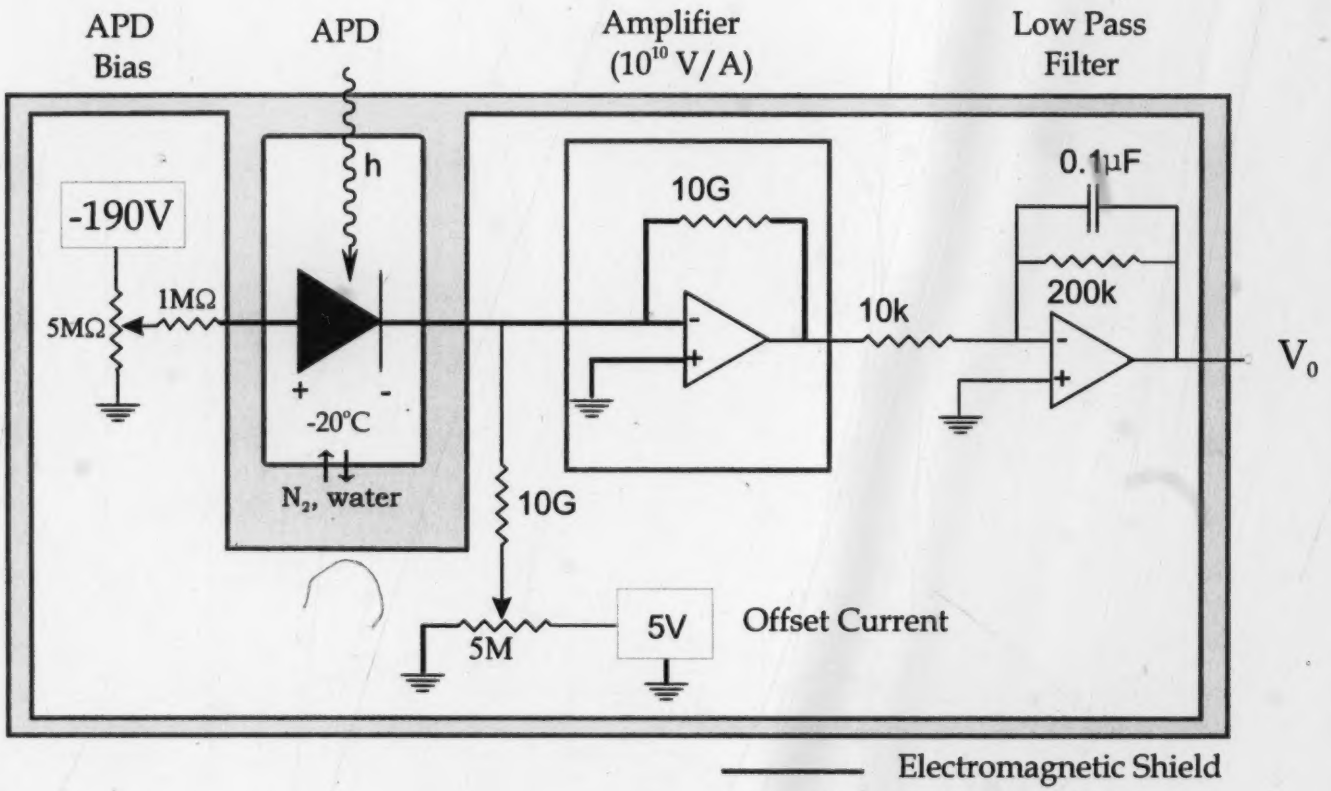


Figure 3.10. APD Electronic Circuitry.

Light incident on APD, biased at -190V, is converted to a current signal. An opposing current offsets the dark current level. The current is converted to voltage with a gain of 10^{10}V/A at the amplifier and low-pass filtered. The APD biases, offset currents, amplifiers and low-pass filters are housed in an aluminum box. The APDs are housed in the copper circuit board box. Both the APD box and the aluminum electronics box are housed in a steel box.

The signal was collected by an analog-to-digital data acquisition board (NB M10-16X, National Instruments Corp., Austin, TX), digitized (16 bit), and recorded in Labview Version 3.1.1 (National Instruments Corp.) on a Macintosh computer (Apple Power Macintosh 7100/66, Apple Computers Inc. Cupertino, CA).

3.2.2.4. Avalanche Photodiode Biasing Voltage

One APD HV power supply (model 522-1, Analog Modules, Inc., Longwood, FL), with eight 5 MΩ pots plus 1 MΩ in parallel, biased eight APDs. The APDs were operated in linear mode and determined to be signal-shot noise limited. The APD biasing voltages were determined by the experiment outlined in Sections 4.2.1.2 and 4.3.1.2.

3.2.3. Capillary Electrophoresis

3.2.3.1. Gel Preparation

DMA Gels for Capillary Refilling

Electroosmosis was reduced in the capillaries with a Grignard coating procedure developed by Novotny [9]. Capillary, 50 µm ID, 150 µm OD (Polymicro Technologies, Phoenix, AZ), was cut into five 3 m lengths. One end of the capillaries was placed in a vial that was pressurized with nitrogen. The nitrogen at a pressure of 20 psi served to force most of the reagents through the capillaries.

The capillaries were rinsed with 1 M NaOH for 1 h, water for 1 h, and methanol for 1 h. The capillaries were placed in an oven at 120°C and flushed overnight with nitrogen at a pressure of 5 psi. The oven temperature was reduced to 65°C and the capillaries were rinsed with thionyl chloride (Sigma, St. Louis, MO) for 30 min. The ends of the capillaries were capped with gas chromatography septa and the capillaries were heated for 8 h at 65°C. The capillaries were washed with 1 mL of 0.3 M vinylmagnesium bromide in freshly distilled anhydrous tetrahydrofuran for 30 min and rinsed with tetrahydrofuran for 30 min. The capillaries were stored with ends capped. The capillaries were washed with water just before *in situ* gel polymerization.

The capillaries were cut into 60 cm lengths and coated with N,N-dimethylacrylamide (DMA, Sigma) as follows. The ends of the capillaries were placed in

a 50-mL 3-neck flask. The capillary coating solution was 4%T DMA in 7 M urea, 400 µL DMA and 4.2 g urea diluted to 10 mL with 1×TTE buffer (100 mM Tris, 100 mM TAPS, 0.1 mM EDTA, pH=8.0). The solution was passed through a 0.22 µm filter into the 3-neck flask and degassed with a vacuum pump for 10 minutes. The polymerization of the DMA was initiated under an argon blanket by adding 40 µL of 10% w/v ammonium persulfate and 20 µL of TEMED (N,N,N',N'-tetramethylethylenediamine). The solution was immediately drawn through the capillaries by vacuum, with the other ends of the capillaries under water. The vacuum was removed and the gel was left to polymerize overnight.

A DMA gel was also prepared for gel refilling, in this case using a 6%T DMA in 7 M urea. The DMA solution was prepared and polymerized as described above. Immediately after adding the TEMED, the solution was drawn into a 10-mL syringe. The syringe end was capped and the gel was allowed to polymerize at room temperature overnight. The gold microtitre plate was filled with the polymerized gel for capillary refilling.

Acrylamide Gels

Capillaries were 50 µm ID, 150 µm OD and 40 cm long. The inner walls of the capillaries were coated using Hjerten's method to covalently bind the polyacrylamide to the walls [10]. A solution of 20 µl of γ-methacryloxypropyltrimethoxysilane (Sigma) in 1 mL of 95% ethanol was drawn through the capillaries for 1 h.

To polymerize the gel in the capillaries, the injection ends were placed in a 50-ml 3-neck flask. The acrylamide solution was 6.5%T in 7M urea, 0.65 g of acrylamide and 4.2 g of urea diluted to 10 ml with 1×TBE buffer (89 mM Tris, 89 mM boric acid and 2 mM EDTA, pH=8.0). The solution was passed through a 0.22 µm filter and added to the flask. The solution was degassed with a vacuum pump for 10 minutes. Under an argon blanket, the polymerization of the acrylamide was initiated by adding 10 µL of 10% w/v ammonium persulfate and 5 µL of TEMED. The solution was immediately drawn through the capillaries by vacuum, with the detector ends of the capillaries under water.

The vacuum was removed and the gel was left to polymerize overnight. Capillaries were visually inspected for bubbles before insertion into the cuvette.

3.2.3.2. DNA Sequencing Samples

The cycle sequencing reactions were prepared using Amersham RPN 2536 kit with ThermoSequenase DNA polymerase (Amersham Life Sciences) and dye-labeled primers (Kit 401487, PE Applied Biosystems, Foster City, CA). The four termination reactions were prepared in four separate tubes. The reagent mix contained the dNTPs, the enzyme, buffer and the appropriate ddNTP. The thermal cycling conditions were 30 cycles of 30 s at 95°C and 30 s at 55°C.

M13mp18 Sample

The single-stranded M13mp18 sequencing sample was prepared with the reagents listed in Table 3.2.

Table 3.2. M13mp18

	A-termination	C-termination	G-termination	T-termination
Reaction ratio A:C:G:T	1	1	1	1
M13mp18	200 ng	200 ng	200 ng	200 ng
M13 forward dye-primer	0.8 pmol (JOE)	0.8 pmol (FAM)	0.8 pmol (TAMRA)	0.8 pmol (ROX)
reagent mix	4 µL	4 µL	4 µL	4 µL
final volume	16 µL	16 µL	16 µL	16 µL

After thermocycling, the 4 reaction tubes were combined and the solution was applied to a Microcon 30 (Amicon, Beverly, MA) membrane. The tubes were spun at 11000 rpm for 4-5 min with the microcentrifuge. The membrane was washed with deionized formamide. The tubes were spun again for 4-5 minutes at 11000 rpm. The tubes were inverted into a new tube and the cleaned DNA spun out of membrane after 4-5 min at 6000 rpm. The final sample volume was about 10 µL.

***Staphylococcus* Samples**

A portion of the Heat Shock Protein-60 kDA gene from *Staphylococcus aureus* was cloned into pUC19. The double-stranded sequencing templates were prepared by PCR across the multiple cloning site with biotinylated primers (Section 5.2.2 to 5.2.4). Table 3.3 lists the recipe for Sample A, a 1:1:1:1 (A:C:G:T) cycle-sequencing reaction. Table 3.4 lists the recipe for Sample B, a 2:2:1:1 (A:C:G:T) cycle-sequencing reaction.

Table 3.3. Sample A

	A-termination	C-termination	G-termination	T-termination
Reaction ratio A:C:G:T	1	1	1	1
PCR template	30 ng	30 ng	30 ng	30 ng
M13 forward dye-primer	0.8 pmol (JOE)	0.8 pmol (FAM)	0.8 pmol (TAMRA)	0.8 pmol (ROX)
reagent mix	4 µL	4 µL	4 µL	4 µL
final volume	16 µL	16 µL	16 µL	16 µL

Table 3.4. Sample B

	A-termination	C-termination	G-termination	T-termination
Reaction ratio A:C:G:T	2	2	1	1
PCR template	60 ng	60 ng	30 ng	30 ng
M13 forward dye-primer	1.6 pmol (JOE)	1.6 pmol (FAM)	0.8 pmol (TAMRA)	0.8 pmol (ROX)
reagent mix	8 µL	8 µL	8 µL	8 µL
final volume	32 µL	32 µL	32 µL	32 µL

After thermocycling, the reaction tubes were combined and the sequencing samples were desalted and cleaned with the following protocol. 15 µL of streptavidin solution immobilized on 4% beaded agarose (Sigma) was added and gently mixed for 20 minutes at room temperature in the dark. The streptavidin fixed the biotinylated PCR products. The agarose beads were pelleted with a quick spin in the microcentrifuge and the supernatant containing dye-labeled, dideoxynucleptide terminated sequencing fragments was applied to a Microcon 30 (Amicon) membrane. The tubes were spun at 11000 rpm for 4-5 min with the microcentrifuge. The membrane was washed with 200 µL of 0.1×TE (1 mM Tris-HCl and 0.1 mM EDTA, pH=8.0). The tubes were spun again

for 4-5 min at 11000 rpm. The tubes were inverted into a new tube and the cleaned DNA spun out of membrane after 4-5 min at 6000 rpm. The final sample volume was about 10 μL.

3.2.3.3. Sample Injection and Electrophoresis

DMA Gels

Sixteen Grignard-DMA coated capillaries were inserted into the sheath-flow cuvette and the reducer fittings in the bomb lid. At the injection end, the capillary tips were placed in a tray filled with buffer, 1×TTE. The buffer was forced through the capillaries at a helium pressure of 250 psi for about 30 min. The tray was replaced with one filled with 6%T DMA gel. At a pressure of 250 psi, the gel was pumped through until it eluted from the other end of the capillaries.

The capillaries were pre-run at 150 V/cm with an aligning solution of 0.4 nM JOE-labeled primer and 0.4 nM of ROX-labeled primer in 1×TTE buffer. The sheath-flow buffer in the cuvette was 1×TTE. The capillaries were either run at room temperature or heated to 40°C by setting the Peltier power supply at a constant current value of 1.5 A. After the alignment of the instrument was confirmed, the aligning solution was migrated from the capillary by electrophoresis of 1×TTE. At the desired temperature, the samples were injected onto the capillaries for 30 s at 100 V/cm. Samples were injected from a gold-plated microtitre plate. The sample injection plate was replaced with a plate filled with 6%T DMA gel and the electrophoresis was continued at 150 V/cm.

Acrylamide gels

Sixteen 6.5%T polyacrylamide gel-filled capillaries were inserted into the sheath-flow cuvette and the reducer fittings in the bomb lid. The capillaries were pre-run with an aligning solution of 0.4 nM JOE-labeled primer and 0.4 nM of ROX-labeled primer in 1×TBE buffer at 150 V/cm. The sheath-flow buffer was 1×TBE. After the alignment of the instrument was confirmed, the aligning solution was then migrated from the capillary by electrophoresis of 1×TBE.

In Run #1, Sample A was injected onto capillary 8 and sample B was injected onto capillary 9. After Run #1 finished, the polarity of the voltage was reversed and the capillaries were run for 1 h at 150 V/cm. In Run #2, sample A was injected onto capillary 8 and Sample B was injected onto capillary 9. The samples were injected for 20 s at 100 V/cm. The sample vials were replaced with a vial filled with 6.5%T polyacrylamide gel and the electrophoresis continued at 200 V/cm and at room temperature.

3.3. RESULTS AND DISCUSSION

3.3.1. Separation Apparatus

3.3.1.1. Capillary Gel Refiller

There are two advantages to gel refilling. First, refilling the capillaries with a gel matrix is an easily automated task compared to replacing the capillaries in the cuvette or pouring new slab gel plates. Second, the DNA separations are more consistent when the capillaries are refilled with gel prepared from a single gel batch. Capillaries prepared by *in situ* gel polymerization varied significantly from batch-to-batch.

The 16-capillary gel refiller was very simple to operate. The HV cable was disconnected from the Plexiglas insulator. A tray filled with the gel matrix was placed on the Teflon holder in the bomb pot. The capillary tips were placed in the gel matrix and the pot was threaded onto the bomb lid. The bomb was then pressurized with gas. These steps took a few minutes to perform.

Two 6%T DMA gels batches were prepared and their viscosities were 500 cp and 1200 cp. The variation in gel viscosity was attributed to the polymerization process. Oxygen was mixed into the solution when it was drawn into the syringe. Oxygen is a free radical scavenger and its presence would result in shorter polymer chains producing a gel with lower viscosity.

The refiller was able to pump low viscosity gel matrix through all 16 capillaries simultaneously. The 500 cp and 1200 cp DMA gels required about 20 min and 60 min respectively to refill 60 cm long, 50 μm ID capillaries at a pressure of 250 psi. In constrast, the polyacrylamide gels *in situ* polymerized for this thesis had viscosities much greater than 5000 cpoise (5 Pa·s).

The refilled polymers were tested for DNA sequencing. The capillaries were refilled with 1200 cp, 6%T DMA gel at room temperature and at 250 psi. The electrophoresis was performed at room temperature and 150 V/cm. A portion of the electropherogram from capillary 16 from file: sb970714-1 is shown in Figure 3.11. The overall capillary current dropped by about 15% (44 μA to 37 μA in about 60 min). The current decay was typically 10% for polyacrylamide sequencing runs.

This initial refilled gel sequencing run was promising. I suspect the poor resolution may be attributed to residual electroosmotic flow at the capillary wall. The electroosmotic flow may arise from incomplete coating of the gel to the capillary wall or the gel coating shearing off during the refilling process. The commercially available ABI DMA polymer is very attractive because the gel itself dynamically coats the capillary wall [2].

In another run, the same capillaries were refilled with the same gel at room temperature and at 250 psi. The capillaries were heated to 40°C and run at 150 V/cm. During the electrophoresis, the current fluctuated and dropped rapidly; the overall current dropped by at least 60% (47 μA to 18 μA in about 70 min). Capillary 11, from file: sb970713-1, is shown in Figure 3.12. Most of the capillaries did not generate peaks. The dramatic current drop suggested that helium dissolution in the gel may be problematic.

According to Henry's Law, the solubility of gas in a liquid is proportional to the partial pressure of the gas on the surface of the liquid. I observed bubbles in the gel tray after refilling the capillaries. Gas solubility also decreases when the temperature of the solution increases. Refilling the capillaries at room temperature and then heating the capillaries will result in the formation of gas bubbles in the gel. In the capillary, bubbles will form, obstruct the flow of ions, and be manifest as a drop in the current. Bubbles were not observed visually when scanning along the capillary length, which was not surprising considering the capillary returned to room temperature.

An idea to fix the solubility problem comes from the design of the Molecular Dynamics' multiple-capillary instrument. Their gel refiller uses nitrogen gas to pressures up 400 psi to refill 75 μm capillaries with hydroxyethylcellulose gel [11]. The entire

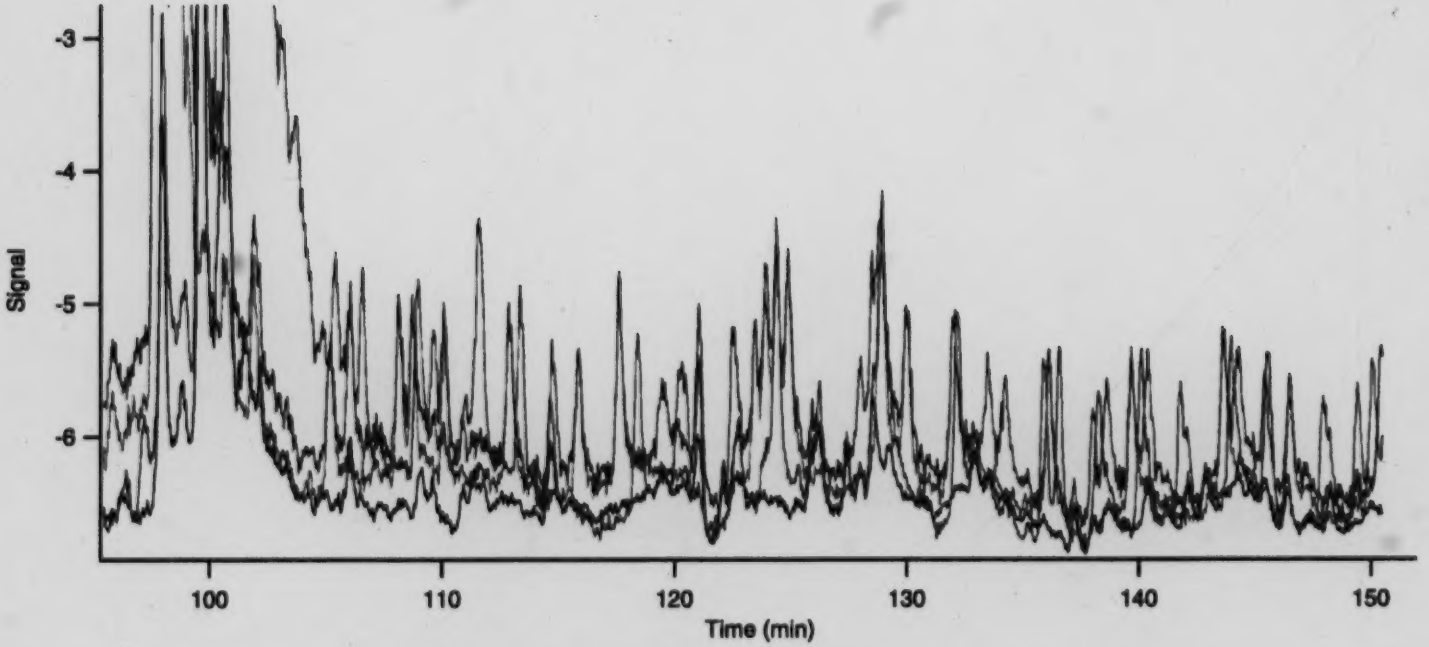


Figure 3.11. DNA Sequencing with Capillary Refilled with DMA, at Room Temperature

The capillary was refilled with 6%T DMA gel at room temperature and helium pressure of 250 psi. The M13mp18 sample was run at 150 V/cm and at room temperature.

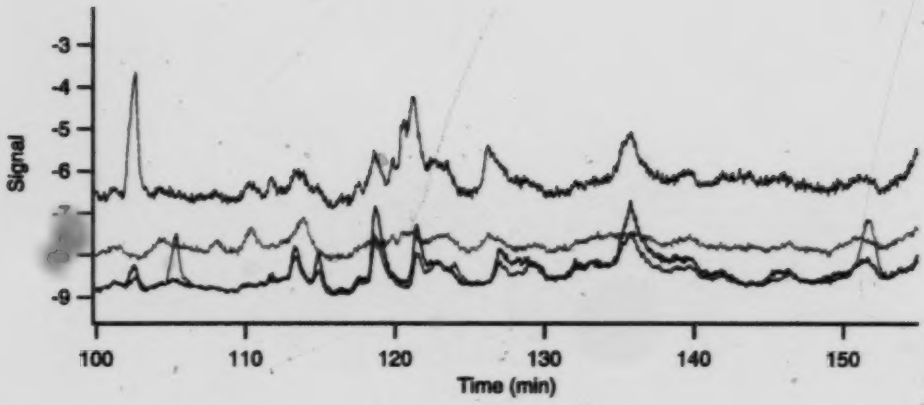


Figure 3.12. DNA Sequencing at 40°C with Capillary Refilled at Room Temperature
The capillary was refilled with 6%T DMA gel at room temperature and a pressure of 250 psi. The M13mp18 sample was run at 150 V/cm and at 40°C. Capillary 11 illustrates the poor quality electropherograms generated. Most capillaries did not have any peaks. The current over the course of the run dropped by 60% in 70 min.

instrument, including the capillary array is thermally equilibrated at 44°C [12] with a precision of ±0.5°C [13]. They report that they can successfully use and refill their capillaries 100 times under these conditions. Perhaps the entire separation apparatus of the 16-capillary instrument, the gel refiller and capillaries, will need to be heated. An advantage of heated refilling will be the decrease in gel viscosity at the higher temperature, decreasing the refill time. However, it may be simpler to build a gel refiller that utilizes liquid pressure. A piston in a cylinder perhaps could mate with a piece that holds the capillaries, similar to the bomb lid, and refill the capillaries [14].

3.3.1.2. Sample Injector

The first generation sample injection method required 30 to 40 min to inject all 16 samples. Four samples, each with an electrode, could be injected at a time. Also the electrical connection of the capillary and electrode was challenging because of the small volume sample (<10 μL). If the electrode and capillary missed the sample solution, which often happened if electrodes bent, the sample was not injected onto the capillary. Thicker electrodes were stiffer, more expensive and sometimes pierced through the bottom of the sample holder during injection.

In a more elegant and simpler design, a gold-plated microtitre plate was both an electrode and a sample holder. Because the entire plate was conducting, the sample solution was always in constant contact with the electrode. Dipping the Pt electrode in a microtitre well filled with buffer ensured the electrical connection to the gold surface. Once the samples were pipetted into the gold microtitre plate, the injection procedure consisted of pulling down the sample carousel, changing the buffer tray to the sample tray, releasing the carousel to mate with Plexiglas insulator, and triggering the HV interlock. The setup and injection of all 16 capillaries required about a minute.

The gold was plated to 300 nm thickness, which dropped to 150 nm on the walls of the wells. Despite the thin layer of gold, there was no noticeable loss of the gold-plating in most of the wells after several runs. The gold-plated microtitre plates were also rinsed with water and ethanol for reuse without loss of gold. However, gold was lost in the well where the platinum electrode made electrical contact. After about 10 runs, the

bottom of the well lost its gold-plating. Instead of sticking the electrode into a gel well, a better method may be to touch the high voltage electrode to the gold surface. The gold-plated microtitre plates were relatively inexpensive to make. There was very little gold, about 0.1g, deposited onto the surface of a 96-well microtitre plate.

Arcing sometimes occurred for this injection device when running at voltages greater than 10 kV. The capillaries were left in the running buffer overnight after a sequencing run, to prevent the gel in the capillary from drying out. The buffer, warm from the electrophoresis, would condense on the steel bomb lid and in the holes of the Plexiglas insulator. The moisture created a leakage pathway and the gold-plated microtitre plate arced to the steel lid of the bomb. Several layers of nail polish were painted on the underside of the bomb lid for electrical insulation.

3.3.1.3. Peltier Capillary Heater

The capillaries were heated with air blowing through a Plexiglas box in the first generation 16-capillary instrument. Temperature regulation was difficult due to the box geometry and the low specific heat capacity of air. A new capillary heater was described by Voss [15]. It was made from Peltier devices sandwiched between two copper plates. Copper has a high specific heat capacity, ensuring constant temperature across the capillary length.

The second generation instrument's Peltier capillary heater was driven with the application of constant DC voltage from the power supply. There was a small drift in the temperature of ±0.5°C over the course of a 3 h run when the system was operated at 45 to 60°C. The drift considered acceptable compared to the temperature oscillations of ±1.5°C that were observed with the forced air heater.

The current settings for capillary heater power supply are listed in Table 3.5.

Table 3.5. Kepco Power Supply Settings for Peltier Heater

Kepco Power Supply Current Setting (A)	Peltier Capillary Heater Temperature (°C)
2.0	45
2.5	55
2.7	60

3.3.2. Detection Apparatus

3.3.2.1. Optics

FAM- and JOE-labeled fragments typically had half the signal of TAMRA- and ROX-labeled fragments. There were two possible methods to improve the signal of these channels; altering the sequencing chemistry or the collection efficiency. The sample could be prepared with higher concentrations of FAM- and JOE-labeled fragments. However, the sequencing dyes are expensive so it was decided to improve the collection efficiency of the instrument.

To increase the signal, the FAM and JOE filters were replaced with 515DF20 and 540DF40 filters respectively. The original filter set consisted of 540DF10, 560DF10, 580DF10 and 610DF10 for the dyes FAM, JOE, TAMRA and ROX respectively. The band pass filters are named for the wavelengths they transmit. For example, a 610DF10 filter transmits a 10 nm bandwidth of light centered at 610 nm, 605 to 615 nm. The narrow 10 nm bandpass on the old filters minimized the spectral overlap in the four channels. The new filters had a wider bandpass, 20 nm and 40 nm, to allow high light transmittance. The new filters were also centered at a shorter wavelength to better match the dyes' emission maximum.

The new bandpass filters and the dye-labeled primers (PE Applied Biosystems) excitation and emission maxima are listed in Table 3.6.

Table 3.6. Fluorescent Dyes and New Bandpass Filters

DNA Base	Fluorescent Dye Primer	λabs, max (nm)	λem, max (nm)	Laser Excitation (nm)	Bandpass filter (nm)
A	FAM	494	521	488 (Ar^+)	515DF20
C	JOE	525	555	488 (Ar^+)	540DF40
G	TAMRA	555	580	532 (Nd:YAG)	580DF10
T	ROX	585	605	532 (Nd:YAG)	610DF10

The two filter sets were compared in Figures 3.13 and 3.14. Run #1 employed the original filter set: 540DF10, 560DF10, 580DF10 and 610DF10 (file: sb970321-1). Run #2 employed the new filter set: 515DF20, 540DF40, 580DF10 and 610DF10 (file:

sb970321-2). Sample A was a 1:1:1:1 mixture of A:C:G:T termination reactions and sample B was a 2:2:1:1 mixture of A:C:G:T termination reactions.

There was a twofold improvement in the signal-to-noise ratio with the new filter set. In Figure 3.13.a, Sample A was detected with the old filter set and with the new filter set in Figure 3.13.b. Run #2 had poorer resolution and a higher background compared to Run #1. The noise was estimated by measuring the standard deviation of the baseline between sequencing peaks. The noise also doubled in Run #2 compared to Run #1 in the FAM and JOE channels. However, there was a fourfold increase in signal in the FAM and JOE channels, C-terminated and A-terminated fragments respectively. The signal and noise from bases 74-77, marked in Figure 3.13, from Sample A are listed in Table 3.7. The signal was in arbitrary units.

Table 3.7. Signal Change in Bases 74-77 Between Run #1 and #2

Run #1	**A**	**C**	**G**	**T**
Signal	3.6	6.6	9.1	17
Noise	0.29	0.31	0.35	0.24
S/N	12	21	26	71
Run #2	A	C	G	T
Signal	17	19	9.0	18
Noise	0.72	0.62	0.28	0.29
S/N	24	31	32	62

Figure 3.14 confirmed increasing the bandwidth of the filters was similar to doubling the sample ratio for the A- and C-terminated fragments. In Figure 3.14.a, Sample B was detected with the old filter set and with the new filter set in Figure 3.14.b. It appeared that the signal was at least four times higher, however some peaks were clipped. Changing to the new filter set increased the signal-to-noise ratio by 50 to 100% for the A and C channels.

3.3.2.2. Avalanche Photodiode Cooling

The first generation version of the APD box (Figure 2.12) was made from Styrofoam to thermally insulate the APD detectors. The new APD box eliminated the

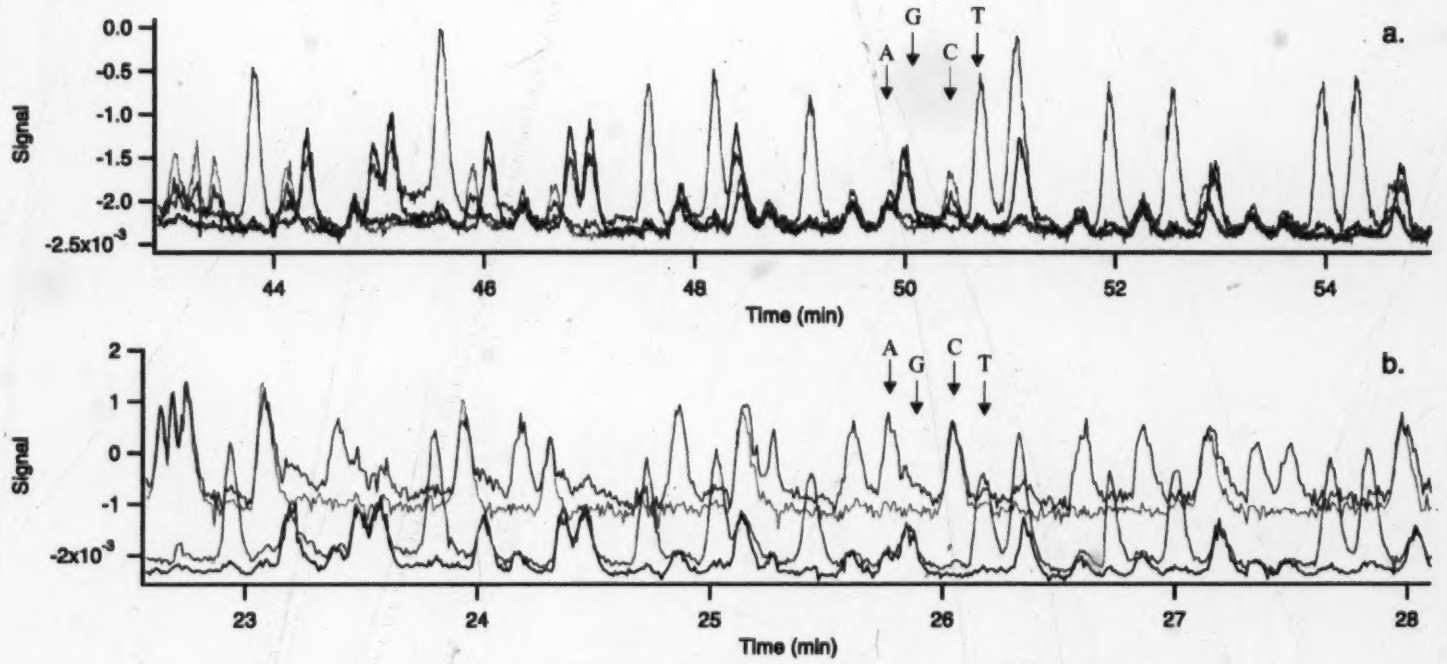


Figure 3.13. Detection of Sample A with Two Different Optical Filter Sets.

A portion of Run #1 and Run #2, 50-90 bases after the forward sequencing primer. The data was unprocessed. The samples were run on 6.5%T polyacrylamide gel, at 200 V/cm and at room temperature. Sample A was a 1:1:1:1 (A:C:G:T) mixture of the termination fragments of *S. aureus*. In a), Sample A was injected onto capillary 8 and detected through the first generation instrument filter set: 540DF10, 560DF10, 580DF10, and 610DF10 (Run #1). In b), Sample A was reinjected onto capillary 8 and detected through the new filter set: 515DF20, 540DF40, 580DF10, and 610DF10 (Run #2).

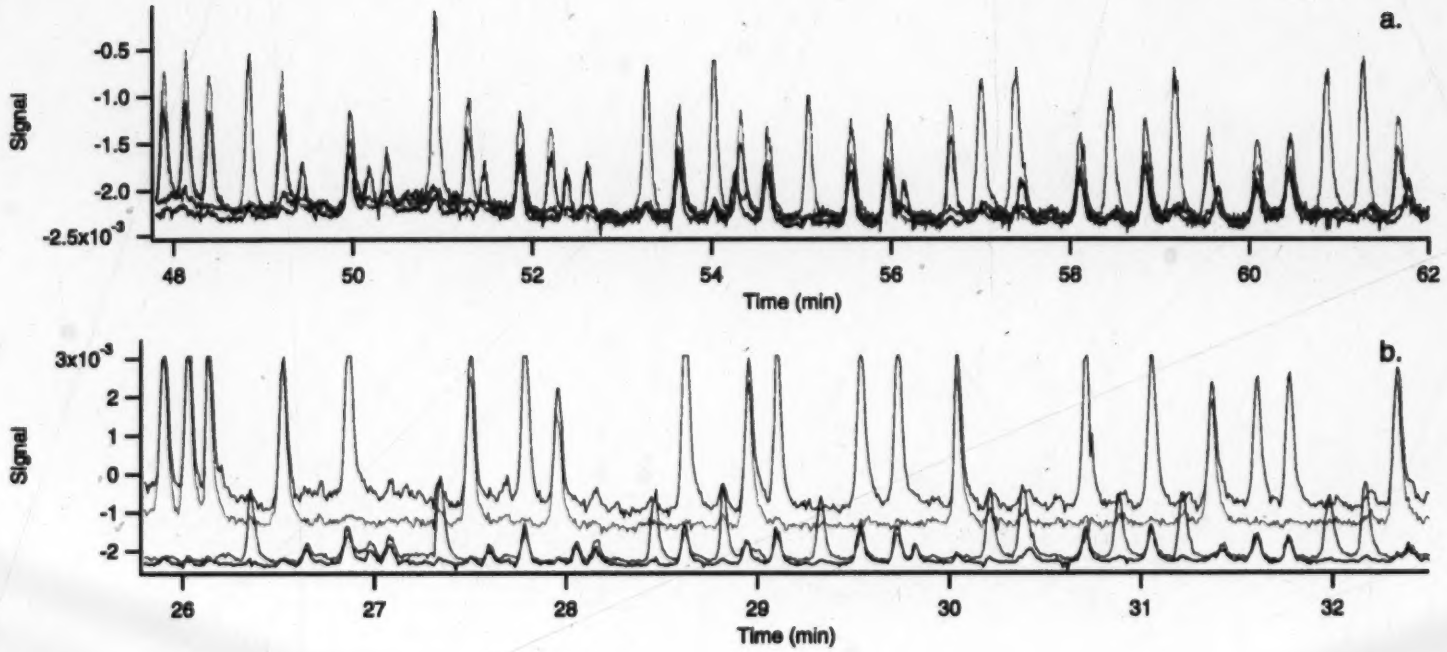


Figure 3.14. Detection of Sample B with Two Different Optical Filter Sets.

A portion of Run #1 and Run #2, 50-90 bases after the forward sequencing primer. The data was unprocessed. The samples were run on 6.5%T polyacrylamide gel, at 200 V/cm and at room temperature. Sample B was a 2:2:1:1 (A:C:G:T) mixture of the termination fragments of *S. aureus*. In a), Sample B was injected onto capillary 9 and detected through the first generation instrument filter set: 540DF10, 560DF10, 580DF10, and 610DF10 (Run #1). In b), Sample B was reinjected onto capillary 9 and detected through the new filter set: 515DF20, 540DF40, 580DF10, and 610DF10 (Run #2).

Styrofoam and replaced it with a simple box constructed from copper-covered circuit-board. The circuit board box successfully maintained the APD temperature at -20°C when cooled with the Peltier device. The copper circuit-board electromagnetically shielded the APDs, replacing the Ad-Mu foil and aluminum foil shields used in the first generation instrument (see Section 2.3.2.3). The controller unit was changed to Proportional-Integral-Derivative (PID) operation. The PID controller reduced the output voltage as it approached the set point temperature. The previous ERO controller simply switched off and on the voltage to the Peltier cooling the APDs (Section 2.2.2.2). Despite trying to optimize the set point and integration time, there remained a low frequency oscillation in the background of approximately 0.01 Hz. Small shifts in the temperature of less than 0.1°C changed the dark current of the APD. These temperature changes generated a few hundred mV drift in the dark signal.

3.3.2.3. Avalanche Photodiode Electronic Circuitry

The noise in the 16-capillary instrument is discussed in Chapter 4. Briefly, there were two dominating sources of noise in the circuitry in the first generation 16-capillary sequencer: background and detector noise. By minimizing the number of components, and by carefully shielding the components, the environmental noise was significantly reduced. Once the interfering noise was reduced, the noise in the APD detectors dominated the background measurement.

After setting the APD biases to the values listed in Table 3.8, the dark current noise in the detectors was 30 mV.

3.3.2.4. Avalanche Photodiode Biasing Voltage

In the first generation instrument, the average bias of the APD was -210 V, and the breakdown voltage around -215 V. In the second generation instrument, the average bias was -195 V and the breakdown around -200 V. It was difficult to compare the biasing voltage values because of the changes made to the instrument. With the high value resistors, the voltmeter used to measure the APD bias changed the resistance of the

circuit. As a result, the voltage readings were different compared to the first generation instrument. The new APD biases are listed in Table 3.8.

Table 3.8. APD Bias at -20.0°C.

APD Number	Old APD Number	New APD Bias (V)
1	4	-180.2
2	3	-192.7
3	2	-194.0
4	1	-187.2
5	8	-191.0
6	7	-195.5
7	6	-194.8
8	5	-189.8
9	12	-175.5
10	11	-196.6
11	10	-196.3
12	9	-194.0
13	16	-188.9
14	15	-185.1
15	14	-192.1
16	13	-198.5

3.4. CONCLUSIONS

This chapter described the improvements made to the 16-capillary instrument.

16 capillaries were simultaneously refilled with low viscosity gel using the capillary gel refiller. The preliminary results, however, from the gel refiller suggested that helium gas dissolved in the gel may ruin the electrophoresis when performed at elevated temperatures.

The sample injection time was tremendously reduced. The spring loaded sample carousel was simple, fast and safe to operate. 16 samples were simultaneously injected with the gold-plated microtitre plate. It can be easily modified to inject many samples, including the injection of 96 samples for the 96-capillary instrument [16].

The collection efficiency was improved for the FAM and JOE channels. The wider bandpass filters improved the signal-to-noise ratio improved about 1.5 to 2 times.

The avalanche photodiode electronic circuitry was rebuilt to improve its signal-to-noise ratio. The gain on the amplifiers was increased, the electronics were

electromagnetically shielded and the APD biasing voltages were optimized. A more detailed description of the methodology to bias the APDs and the limit of detection of the 16-capillary instrument can be found in the Chapter 4.

In Chapter 5, the second generation instrument is used to sequence *Staphylococcus* samples.

3.5. REFERENCES

(1) ABI PRISM 310 Genetic Analyzer, Perkin-Elmer/Applied Biosystems, **1997**.

(2) Rosenblum, B. B.; Oaks, F.; Menchen, S.; Johnson, B. *Nucleic Acids Research* **1997,** *25*, 3925-3929.

(3) *CRC Handbook of Chemistry and Physics*, 63 ed.; CRC Press Inc.: Boca Raton, Florida, **1982.**

(4) Voss, K. O., personal communication, **1997**.

(5) Barton, A. F. M. *CRC Handbook of Solubility Parameters and other Cohesion Parameters*; CRC Press, Inc.: Boca Raton, Florida, **1983**.

(6) Gerrard, W. *Solubility of Gases and Liquids: A Graphical Approach*; Plenum Press: New York, **1976**.

(7) Polonsky, G., Strength calculations for pressure vessel "Mini-Bomb", Lawrence Berkeley National Laboratory, **1996**.

(8) Lewis, D. F., Ph.D. Thesis, in preparation, University of Alberta, .

(9) Cobb, K. A.; Dolnik, V.; Novotny, M. *Analytical Chemistry* **1990,** *62*, 2478-2483.

(10) Hjerten, S. *Journal of Chromatography* **1985,** *347*, 181-198.

(11) Bashkin, J.; Marsh, M.; Barker, D.; Johnston, R. *Applied and Theoretical Electrophoresis* **1996,** *6*, 23-28.

(12) Dolnik, V., personal communication, Molecular Dynamics, **1998**.

(13) Bashkin, J. S.; Bartosiewicz, M.; Roach, D.; Leong, J.; Barker, D.; Johnston, R. *Journal of Capillary Electrophoresis* **1996,** *3*, 61-68.

(14) Dovichi, N. J., personal communication,**1997**.

(15) Voss, K. O. Ph.D., University of Alberta, Edmonton, Alberta, **1998**.

(16) Zhang, J. Z. Z.; Dovichi, N. J., in preparation.

Chapter 4: Optimizing the Avalanche Photodiodes

4.1. INTRODUCTION

The16-capillary sequencer was rebuilt to improve the signal-to-noise ratio (S/N) of the avalanche photodiode (APD) detectors. The sensitivity of an APD increased with increasing bias as the *internal* gain of the APD increased. However, the signal-to-noise ratio decreased at higher biases as the excess noise in the APD dominated the system. To increase the signal without increasing the noise, the *external* gain of an APD was increased. Described in Section 3.2.2, the second generation instrument had higher external gain transimpedance amplifiers and better shielding to reduce environmental noise.

The APD's signal-to-noise ratios were compared to a Single Photon Counting Module (SPCM). The SPCM, from the 5-capillary instrument, is a Geiger mode APD and is very sensitive. Zhang reported a limit of detection (LOD) of 130 molecules of fluorescein dye on the 5-capillary instrument [1].

This chapter outlines the procedure used to enhance the performance of the 16-capillary instrument's APD detectors. First, the noise in the APDs was characterized. Second, the signal-to-noise ratios of the APDs were compared to the SPCMs from the 5-capillary instrument to determine the appropriate APD bias. Third, the collection optics between the two instruments were compared. Finally, a limit of detection for the 16-capillary instrument was determined on the optimized APDs.

4.1.1. Noise

Quantum noise is a fundamental noise arising from random emission of the photons from a source. Shot noise is a specific name for fundamental noise at the signal detectors arising from the random conversion of photons to signal. The Poisson distribution describes the standard deviation in the number of photons, σ_{shot}, detected in time period t:

$$\sigma_{shot} = (\text{photons} / \text{sec} \cdot t)^{1/2} = (n)^{1/2} \qquad (4.1)$$

where n is the mean number of photons counted in time t. Another fundamental noise, Johnson ("thermal") noise, originates from the random thermal motion of electrons in resistors and is described as:

$$\sigma_j = (4kTR\Delta f)^{1/2} \tag{4.2}$$

where σ_J is the Johnson noise, k is Boltzmann's constant, T is the temperature in Kelvin, R is the resistance, and Δf is the frequency bandwidth of the system.

Nonfundamental noise is derived from imperfections in instrumentation and can be theoretically eliminated. Environmental noise arises from the instrument coupling to other signal sources such as the 60 Hz noise from AC power lines. Flicker noise describes the noise that is directly proportional to signal. 1/f noise, a form of flicker noise, is inversely proportional to frequency; the noise has greater magnitudes at low frequencies from a slow drift in the DC component of the signal. Flicker noise is observed experimentally and it may be derived from non-60 Hz noise in the electrical power supply, fluctuations in the light source, or from mechanical vibrations. Quantization noise arises from the finite resolution of a readout device. It occurs when an infinite resolution signal is converted to a set of finite numbers in the analog-to-digital conversion process.

Noise arises from several primary sources: the analytical signal, the background, and the detector. The total noise σ_{total} can be described as [2]:

$$\sigma_{total} = \left(\sigma_S^2 + \sigma_B^2 + \sigma_D^2\right)^{1/2} \tag{4.3}$$

$$\sigma_S = \left(\sigma_{S,shot}^2 + \sigma_{S,flick}^2\right)^{1/2} \tag{4.4}$$

$$\sigma_B = \left(\sigma_{B,shot}^2 + \sigma_{B,flick}^2\right)^{1/2} \tag{4.5}$$

$$\sigma_D = \left(\sigma_{Dark}^2 + \sigma_{ar}^2\right)^{1/2} = \left(\left(\sigma_{Dark,shot}^2 + \sigma_{Dark,exc}^2\right)^{1/2} + \sigma_{ar}^2\right)^{1/2} \tag{4.6}$$

where σ_S is the noise in the analytical signal, σ_B is the noise in the background, σ_D is the noise in the detector, $\sigma_{S,shot}$ and $\sigma_{S,flick}$ are the shot and flicker noise in the signal respectively, $\sigma_{B,shot}$ and $\sigma_{B,flick}$ are the shot and flicker noise in the background signal respectively, σ_{Dark} is the total dark current noise, $\sigma_{Dark,shot}$ is the dark current shot noise and $\sigma_{Dark,exc}$ is the excess dark current noise arising from nonthermal sources. σ_{ar} is the amplifier/readout noise and observed in the absence of the analytical signal, background, and dark current. The signal-to-noise (S/N) expression is:

$$\frac{S}{N} = \frac{S}{\sigma_{total}} \tag{4.7}$$

where $S = S_O - (S_B + S_D)$, S is the analytical signal, S_O is the signal output from the detector, S_B is the background signal, and S_D is the detector signal in the absence of the signal and the background.

Noise is often dominated by one source. Ideally, the system is shot-noise-limited. In a background-shot-noise-limited situation, the S/N can be improved by decreasing the background. If the noise is broadband, such as shot and Johnson noise, limiting the frequency bandwidth will reduce the noise. In a dark-current-shot-noise-limited situation, the S/N can be improved by increasing the signal or by reducing the dark current levels by cooling the detectors.

4.1.2. Avalanche Photodiodes

The output signal current, I_S, from an APD is [3]:

$$I_S = MR_0(\lambda)P_{rad} \tag{4.8}$$

where M is the average internal gain that is a function of biasing voltage and P_{rad} is the incident radiant power. $R_0(\lambda)$ is the intrinsic responsivity (in A/W) of the APD at M=1 and wavelength λ and $R_0(\lambda)$ is directly proportional to the APD's quantum efficiency.

At low light levels, the APD noise is dominated by the dark current shot noise, $\sigma_{Dark,Shot}$. The APD signal and noise expressions at low light levels are [3, 4]:

$$I_{Dark} = I_{DS} + I_{DB}M \tag{4.9}$$

$$\sigma_{Dark,shot} = \left[2q\Delta f\left(I_{DS} + I_{DB}M^2F\right)\right]^{1/2} \tag{4.10}$$

where I_{Dark} is the total dark current signal, I_{DS} is the surface dark current, I_{DB} is the bulk dark current, M is the internal gain, q is the electron charge, and Δf is the bandwidth of the system. F is an excess noise factor because the APD is not a perfectly noiseless multiplier due to the statistical nature of the avalanche process; not all the photogenerated carriers have the same gain. F is given by [5]:

$$F = kM + (1-k)\left(2 - \frac{1}{M}\right) \tag{4.11}$$

where the ionization ratio $k = \beta/\alpha$, the hole-to-electron ionization rates ratio ($k \leq 1$). The ionization ratio is a function of electric field across APD, and in silicon, k is lowest at low electric fields. At high electric fields, k increases as does M. Thus the excess noise factor increases with bias. In reach-through structure silicon APDs, k is about 0.02 [3].

At low gain (10^2), the gain can be written as a function of biasing voltage [4]:

$$M = \frac{I}{I_{DS}} = \left[1 - \left(\frac{V_J}{V_{BR}}\right)^n\right]^{-1} \tag{4.12}$$

where I is the multiplied output current from APD, V_J is the junction voltage, V_{BR} is the breakdown voltage and n is a factor depending on diode structure and material.

4.1.3. Limit of Detection

Conventionally, the limit of detection is defined as the analyte concentration that yields a net analyte signal (S) equal to *m* times the standard deviation (σ_B) of the background (B):

$$C_L = m\sigma_B \frac{C_i}{S_i} \tag{4.14}$$

where C_L is the concentration detection limit, and C_i is concentration that creates a net signal S_i. The value *m* is usually equal to 3 [6].

Mass detection limits are estimated by calculating the volume of sample injected on the capillary, Vol_{inj} [7]:

$$Vol_{inj} = Vol_{cap} \times \frac{t_{inj}}{t_{mig}} \times \frac{V_{inj}}{V_{sep}} \tag{4.15}$$

where Vol_{cap} is the volume of the capillary, t_{inj} is the time sample is injected, t_{mig} is the migration time of sample peak, and V_{inj} and V_{sep} are the voltage for injection and separation respectively.

4.2. EXPERIMENTAL

4.2.1. Noise

4.2.1.1. Noise Intensity *vs.* Frequency at Increasing APD Bias

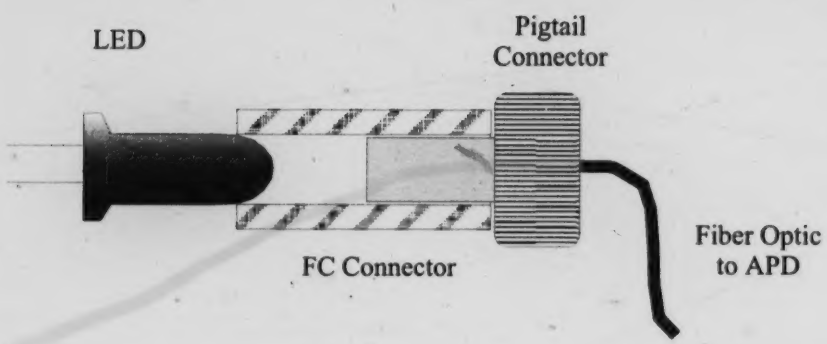
APD #8 (C30902S, EG&G Canada, Vaudreuil, Quebec) was used for these measurements. The APD output was collected in the dark with the fiber optic ends capped. The APD output was conditioned by a transimpedance amplifier (model 341, Analog Modules, Inc., Longwood, FL) with 10^{10} V/A gain. The signal was not low-pass filtered. The APD bias (HV Power Supply, model 522-1, Analog Modules, Inc.) was increased from -140 V to -194 V. One set of measurements were acquired with the steel box lid off (the outermost electromagnetic shield) and one set of measurements were acquired with the box lid on. A 512 point, 1250 Hz range Fourier transform of the APD dark current signal was collected by an oscilloscope (54542A, Hewlett Packard Company, Colorado Springs, CO).

4.2.1.2. Signal-to-Noise Studies with Increasing APD Bias and External Gain

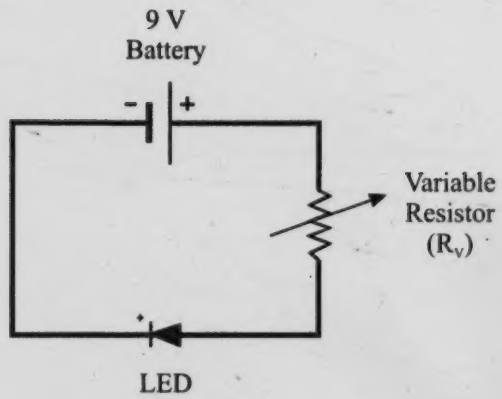
A green light-emitting diode (LED, Knight Lites) was used as a light source in these experiments (Figure 4.1.a). An APD-fiber optic cable (50 μm core/125 μm cladding) was connected to an FC connector where it mated with the green LED. In Figure 4.1.b, varying the voltage dropped across the LED with a variable resistance device (Ohm Ranger, Ohmite Manufacturing Company, Skokie, IL) attenuated the LED light intensity. The APD output was externally amplified with a gain of 10^9 V/A and conditioned with an 8 Hz low-pass filter. The measurements were repeated at an external gain of 10^{10} V/A and the same low-pass filter.

Similar measurements were acquired with an SPCM detector (SPCM-100-PQ, EG&G Canada). A GRIN-fiber optic (50 μm core/125 μm cladding) was connected to the SPCM. The GRIN lens fit into the FC connector with the LED. The SPCM output was connected to a Frequency-to-Voltage Converter (Larry Coulson, March 1994).

The APD and SPCM outputs were read off the oscilloscope display.



a.



b.

Figure 4.1. The LED in the Signal-to-Noise Experiments.
In a), the LED is glued to an FC connector. A fiber optic pigtail connector transmits the LED light to the APD. In b), an electronic circuit diagram of the LED circuit is shown. The voltage from the battery is adjusted with a variable resistor R_v attenuating the current through the LED. The light intensity is inversely proportional to the resistance in the circuit.

4.2.2. Limit of Detection

4.2.2.1. Fiber Optic Core Comparison on 5-Capillary Instrument

A 42.5 cm, 50 μm ID capillary (Polymicro Technologies, Phoenix, AZ) was used for the separation on the 5-capillary instrument. The running buffer and sample buffer were 40 mM borate, 10 mM SDS, pH = 9.2. Solutions of 1×10^{-9} and 2×10^{-10} M fluorescein (BDH Chemicals, Poole, England) were injected onto the capillary for 5 s at 4 kV and run at 15 kV. A blank was run by dipping the end of the capillary in the sample for 5 s without applying the voltage and run at 15 kV (350 V/cm). The injection volume of the sample was about 3 nL. Subtracting the peak height of the siphon peak (the blank run) from the injected peak compensated sample siphoning. The Ar^+ laser power was 5.8 mW.

The fluorescence from the capillaries was collected by a 20 ×, 0.5 NA microscope objective (Melles Griot). The signal was passed through a 540DF10 bandpass filter (Omega Optical) and imaged onto a 1-mm diameter GRIN lens (SELFOC® Fiber Collimators, FCM-00F-050-0.63, Nippon Sheet Glass Co., Tokyo, Japan) connected to a fiber optic cable (100 μm core/140 μm cladding). The fiber optic cable connected to SPCM. The sample injections were repeated and collected with a 1-mm diameter GRIN lens connected to a smaller core fiber optic cable (50 μm core/125 μm cladding).

The signal was converted to voltage, collected by an analog-to-digital data acquisition board (NB-IO-10, National Instruments Corp., Austin, TX), digitized (16 bit) and recorded in Labview (National Instruments Corp.) on a Macintosh computer (Apple Computers Inc. Cupertino, CA). The data were smoothed using Savitzsky-Golay second order filter of 21 points.

4.2.2.2. LOD Measurements on 16-Capillary Instrument

A 42.5 cm, 50 μm ID capillary was used for the separation on the 16-capillary instrument. A fluorescein solution of concentration 2×10^{-11} M was injected onto the capillary as described in Section 4.2.2.1. The Ar^+ laser power was 30 mW, measured at the laser head.

The fluorescence from the capillaries was collected by a 18 ×, 0.45 NA microscope objective (Melles Griot) and passed through 540DF10 bandpass filter (Omega Optical). The signal was imaged onto a 1-mm diameter GRIN lens connected to a fiber optic cable (50 µm core/125 µm cladding). The fiber optic coupled at an FC connector to another fiber optic cable connected to APD #5. The second fiber optic had the same diameter core. The APD was biased at -191.0 V. The APD output was amplified with a gain of 10^{10} V/A and conditioned with an 8 Hz low-pass filter.

The signal was collected by an analog-to-digital data acquisition board (NB M10-16X, National Instruments Corp.), digitized (16 bit) and recorded in Labview Version 3.1.1 (National Instruments Corp.) on a Macintosh computer (Apple Computers Inc. Cupertino, CA).

4.3. RESULTS AND DISCUSSION

4.3.1. Noise

4.3.1.1. Noise Intensity *vs.* Frequency at Increasing APD Bias

The Fourier transforms of the dark current at various APD biases are shown in Figures 4.2 and 4.3. In Figure 4.2, the measurements were collected with the steel electronics box lid off. In Figure 4.3, the measurements were acquired with the steel electronics box lid on.

A change in the noise-limiting situation is illustrated by Figures 4.2 and 4.3. At low APD bias, low internal gain, the APD was dominated by background noise. At high APD bias, high internal gain, the APD noise dominated over background noise. Less noise was observed at the lower APD biases of -140 V to -190.1 V. The quantization noise from the oscilloscope was seen from 600 to 1000 Hz. In this region, the noise was smaller than 0.01 mV, which is the analog-to-digital resolution of the oscilloscope (20 V/2^{11} bit = 0.01 mV/bit). At biases greater than -190.1 V, the noise was larger than 0.01 mV. The quantization noise was not observed because it was dominated by other noises in the system. The change in the noise-limiting situation was also demonstrated in Figure 4.2. At the first 6 APD biases, -140.0 to -191.1 V, the 60 Hz environmental peak

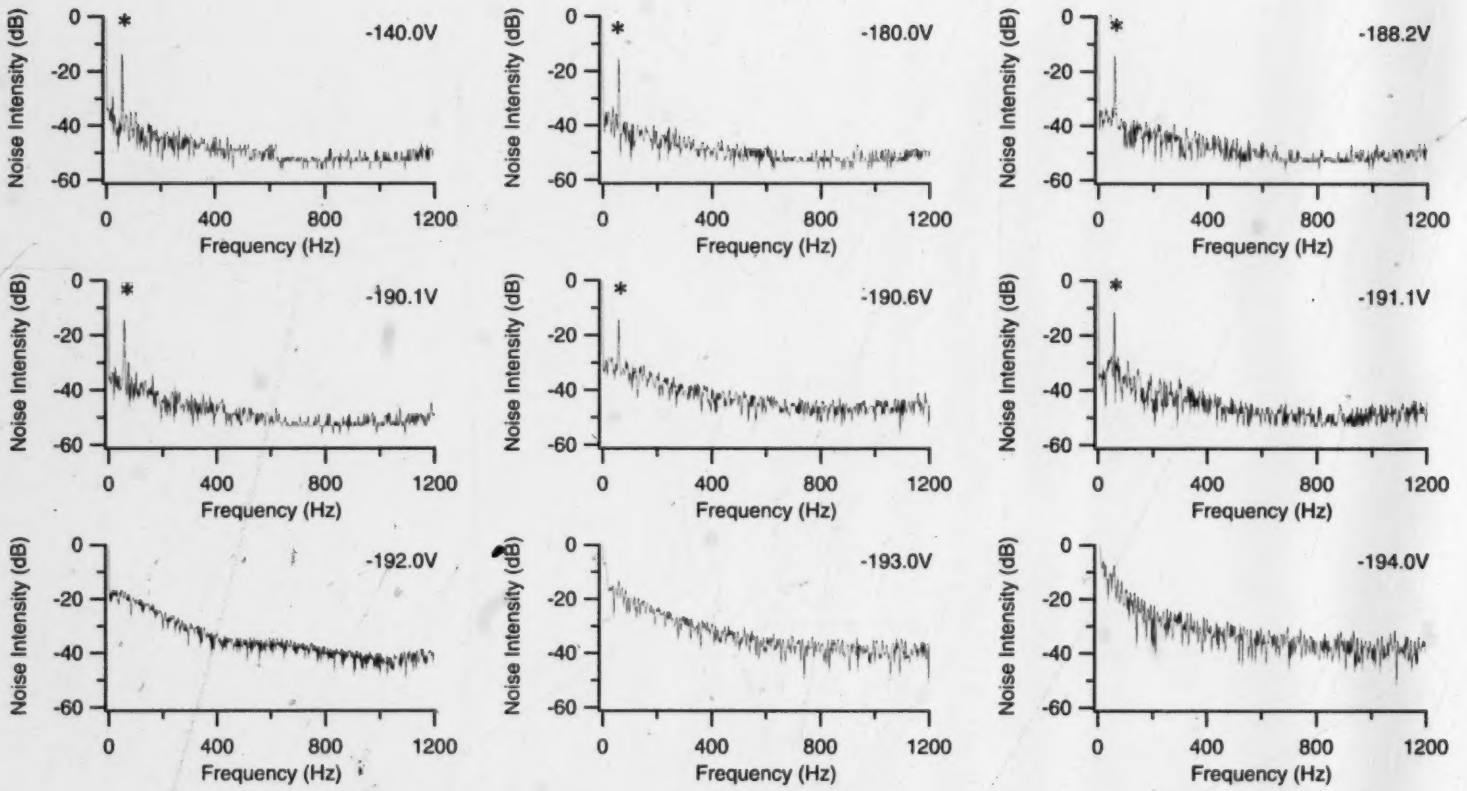


Figure 4.2. Dark Current *vs.* Frequency of APD Output at Increasing Bias.
With the steel box lid removed, the 60 Hz environmental noise, marked with an asterisk (*) can be seen at APD biases -140.0V to -191.1 V. The APD external gain was 10^{10}V/A and the APD output was not low-pass filtered.

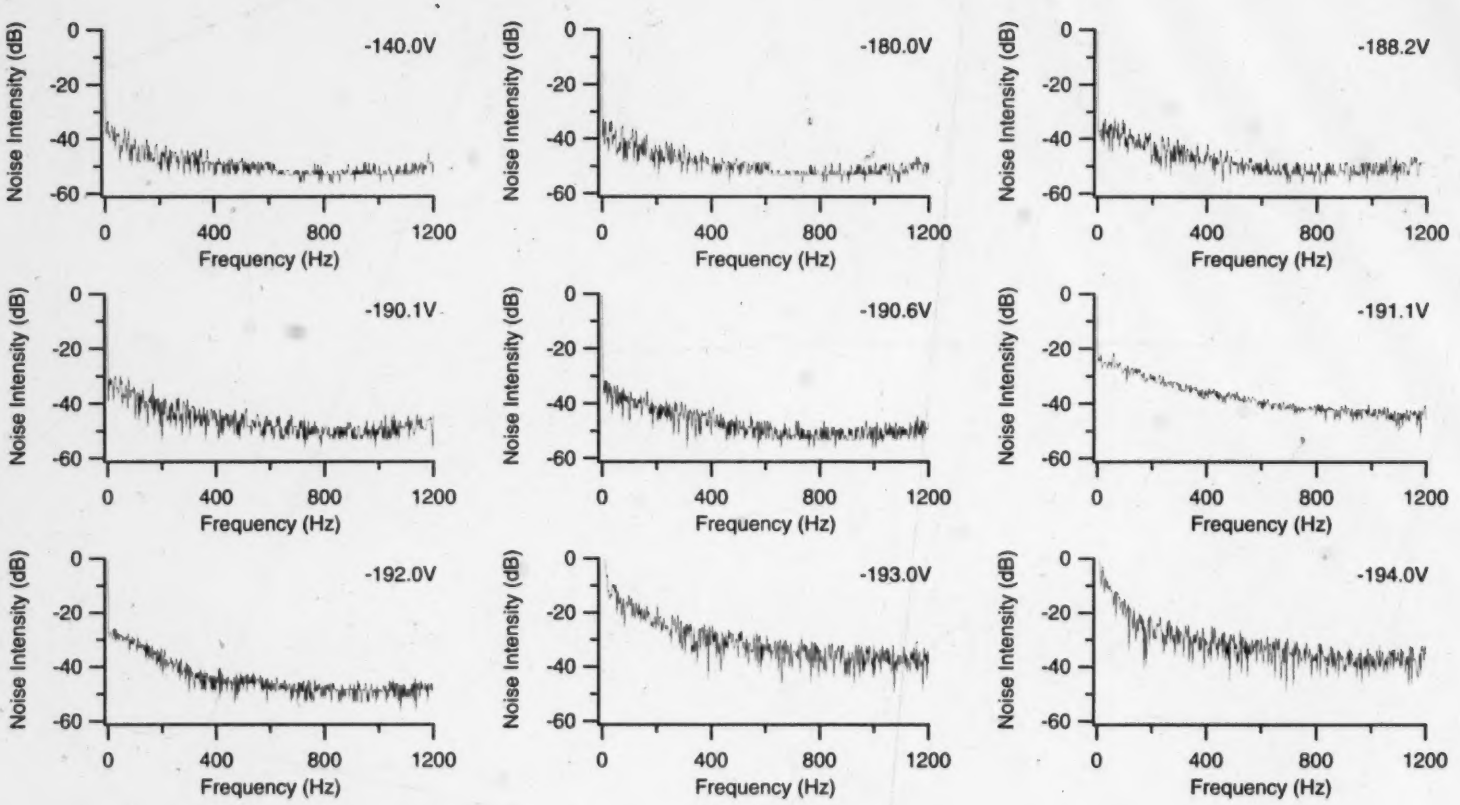


Figure 4.3. Dark Current *vs.* Frequency of APD Output at Increasing Bias.
With the steel box lid on, the electronics are shielded from the 60 Hz environmental noise. The APD external gain was 10^{10}V/A and the APD output was not low-pass filtered.

was conspicuous. However, at biases greater than -192 V, the APD noise obscured the 60 Hz peak.

The increase in the APD noise with increasing bias could be seen in the time domain on the oscilloscope. Spurious peaks in the APD output signal were observed as the bias approached the breakdown voltage of the APD at -194 V. The frequency of these peaks increased with APD bias; eventually the dark current saturated the oscilloscope input at the +10 V rail.

The noise intensity exponentially decayed with frequency. This relationship was attributed to the transimpedance amplifiers. The amplifiers have approximately 200 Hz cutoff frequency at gain of 10^{10} V/A. The noise intensity decay may also be the result of flicker noise with a 1/f component; its noise intensity is greater at lower frequencies. I observed a drift in the baseline of the APD signal, which I attributed to small changes in the temperature affecting the APD dark current.

In Figure 4.3 the measurements were acquired with the steel lid of the electronics box on. The absence of the 60 Hz noise peak demonstrated the electromagnetic shielding of the electronics by the steel box.

4.3.1.2. Signal-to-Noise Studies with Increasing APD Bias and External Gain

The LED light level was inversely proportional to the resistance in the circuit in Figure 4.1. In the following discussion, the LED light level was expressed as $1/R_V$ in units of $k\Omega^{-1}$, where R_V is the value of the variable resistor. For example, the LED light intensity was 0.01 for a resistor value of 100 kΩ.

Increasing the APD bias increased its internal gain but also increased the noise in the APD. Figure 4.4.a demonstrates this trend as the APD signal increased with increasing bias. Figure 4.4.b demonstrates that the signal-to-noise ratio decreased with increasing bias.

The effect of bias voltage on the signal and signal-to-noise ratios was further investigated. In Figure 4.5.a, the APD signal is plotted as a function of light intensity at four APD biases. In Figure 4.5.b, the APD signal-to-noise ratio is plotted as a function of light intensity at four APD biases. The APD had an external gain of 10^9 V/A, which was

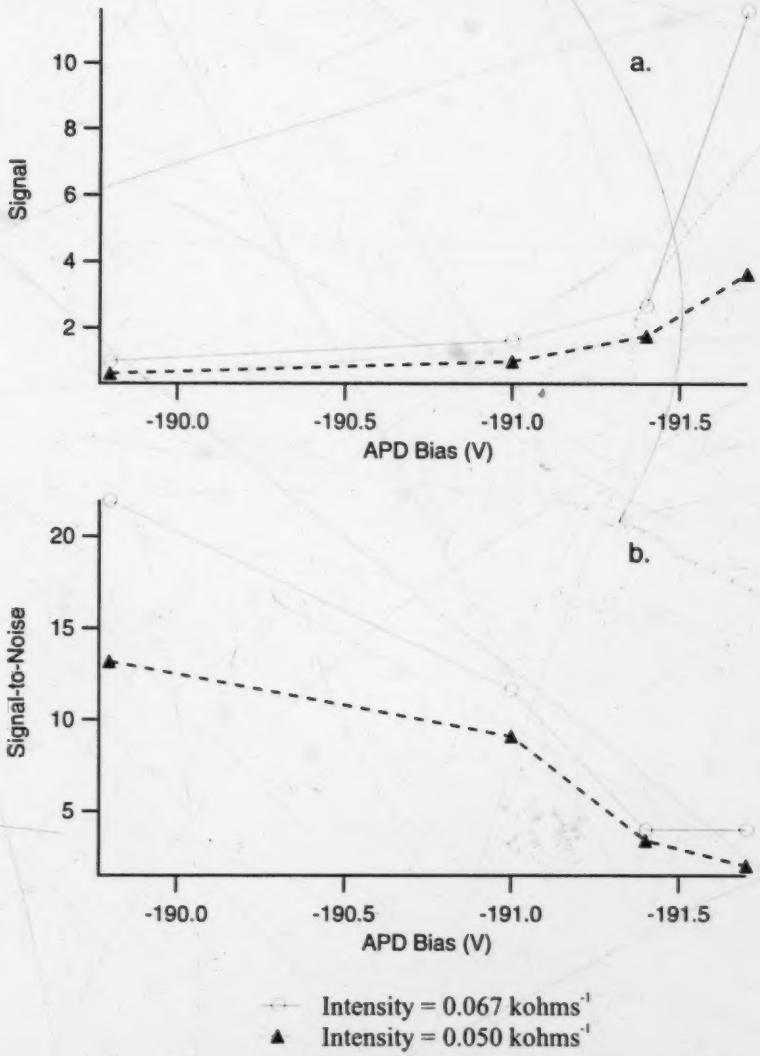


Figure 4.4. APD Response *vs.* APD Bias at External Gain of 10^9V/A
In a), the APD signal is plotted as a function of APD bias at two light intensities. In b), the signal-to-noise ratio is plotted as a function of APD bias at two light intensities.

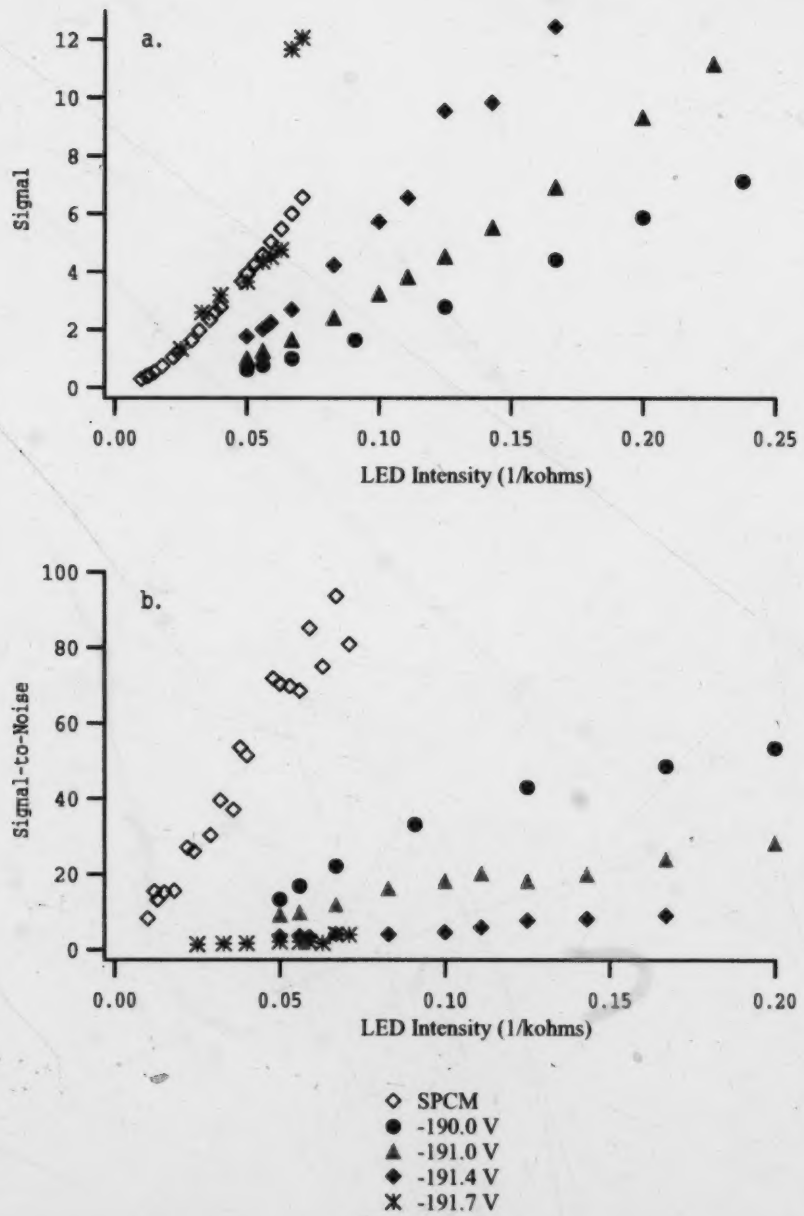


Figure 4.5. APD Response to LED Light Intensity at APD External Gain 10^5V/A. In a), the APD output signal is plotted *versus* Intensity. In b), the signal-to-noise ratio of the APD is plotted *versus*. Intensity (expanded). The intensity of the LED is inversely proportional to the resistance. In a), at a bias of -191.7 V, two points deviate from linearity around 0.07 kohms^{-1}.

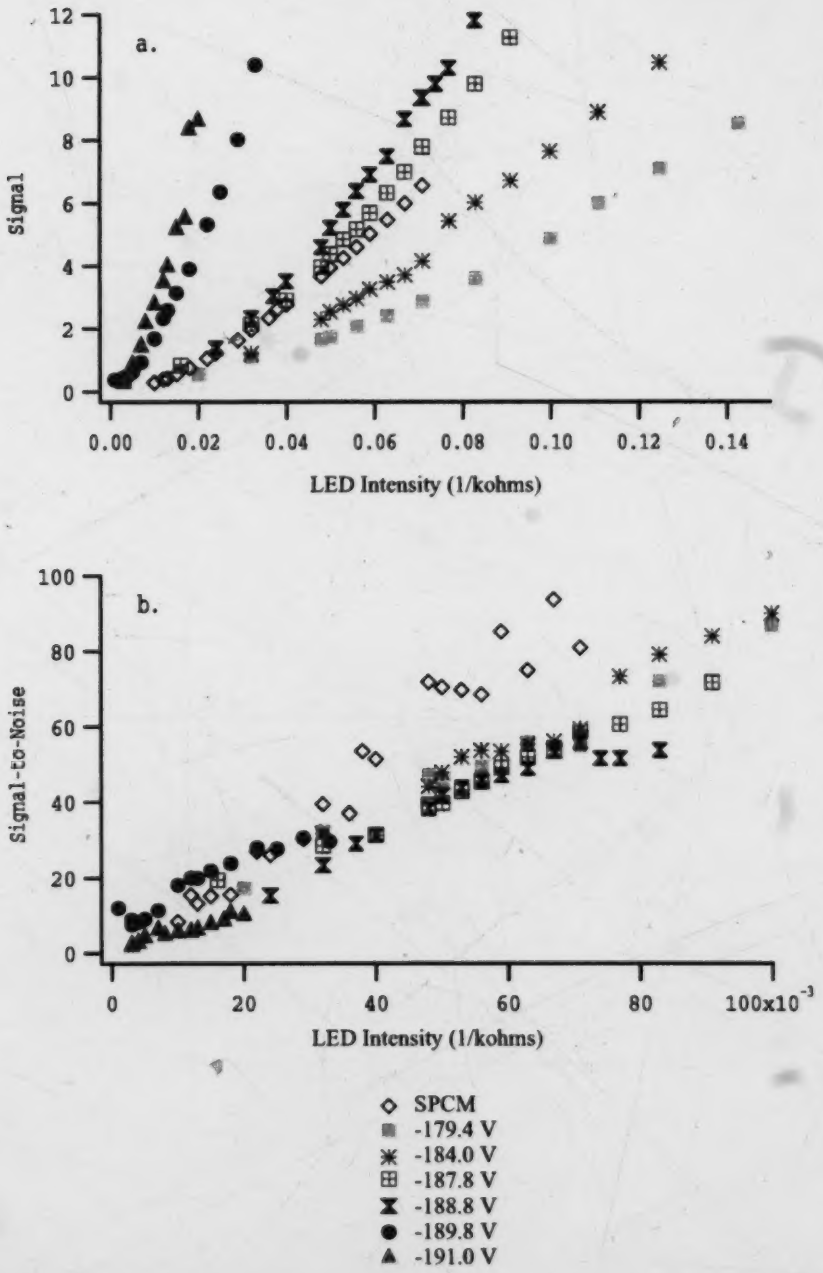


Figure 4.6. APD Response to LED Light Intensity at APD External Gain 10^{10}V/A.

In a), the APD output signal is plotted *versus* Intensity. In b), the signal-to-noise ratio of the APD is plotted *versus* Intensity (expanded). The intensity of the LED is inversely proportional to the resistance. Note the intensity scale is changed compared to Figure 4.5.

used in the first generation instrument. The sensitivity of the APDs, the slope of the signal *vs.* light intensity plot, increased as the bias increased from -190.0 V to -191.7 V (Figure 4.5.a). However, this increase in sensitivity was marred by the decrease in the signal-to-noise ratio (Figure 4.5.b). Furthermore, the signal-to-noise ratios of the APDs were much lower than the SPCM values. The decrease in signal-to-noise with the increased *internal* gain demonstrated that the noise in the detector was dominating the system.

It was hoped that increasing the *external* gain of the detectors would improve the signal-to-noise ratio. A lower APD internal gain could be employed with a higher APD external gain. Also, if the capacitance is held constant in the amplifier feedback loop, increasing the resistance would increase the time constant, decreasing the bandwidth of the amplifier. Decreasing the bandwidth would serve to low-pass filter the APD output.

The effect of the APD response at an external gain of 10^{10} V/A gain is presented in Figure 4.6. In Figure 4.6.a, the APD signal is plotted as a function of light intensity. As before, the sensitivity of the APDs, the slope of the signal *vs.* LED light intensity plot, increased as the bias increased. In Figure 4.6.b, the APD signal-to-noise ratios were plotted as a function of light intensity. In this figure, the APD signal-to-noise ratios were similar to the SPCM. The APD performance was improved by increasing the external gain to 10^{10}V/A.

What was the optimal APD bias? At low APD bias, the APD was dominated by background noise; it had difficulty detecting low light levels. Increasing the bias, increased the internal gain, and increased the APD sensitivity. But at very high APD bias, the APD was dominated by detector noise. At constant light intensity, a plot of signal-to-noise as a function of APD bias should reach a maximum at the optimal bias. However, it was difficult to find a light intensity that could be detected at low APD bias and that did not saturate the oscilloscope range at the high bias. To extrapolate a common light intensity, the relationship between signal-to-noise and light intensity was calculated using a linear approximation analysis.

The linear approximation calculation had one small complication. In this chapter, the LED light intensity was defined as $1/R_V$, the variable resistor value in the circuit

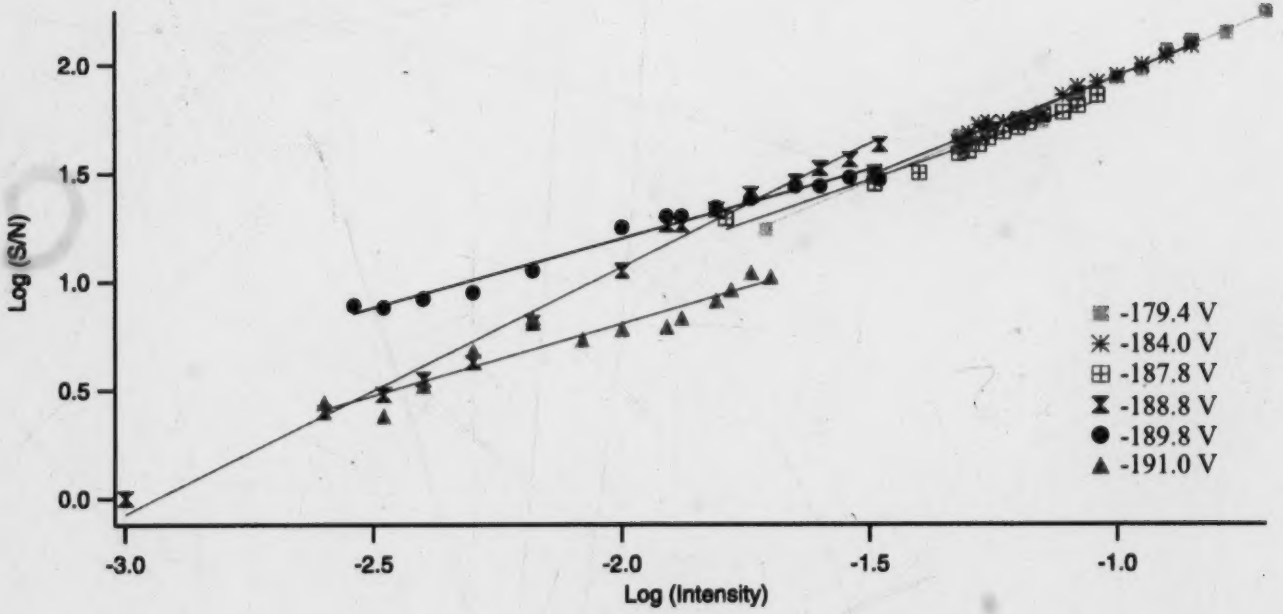


Figure 4.7. Log Signal-to-Noise *vs.* Log Intensity at External Gain of 10^{10} V/A

At -179.4 V, log (S/N) = 0.97 log (Intensity) + 2.9
At -184.0 V, log (S/N) = 0.93 log (Intensity) + 2.9
At -187.8 V, log (S/N) = 0.79 log (Intensity) + 2.7
At -188.8 V, log (S/N) = 1.10 log (Intensity) + 3.3
At -189.8 V, log (S/N) = 0.64 log (Intensity) + 2.5
At -191.0 V, log (S/N) = 0.66 log (Intensity) + 2.1

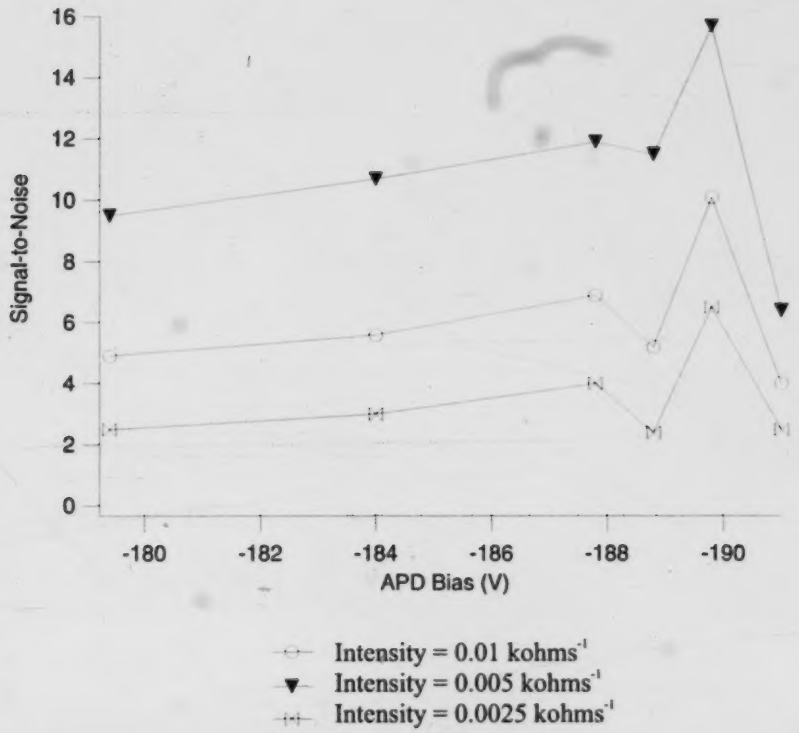


Figure 4.8. Signal-to-Noise Ratio *vs.* APD Bias at External Gain of 10^{10}V/A
The APD signal-to-noise ratios were calculated from the linear approximations in Figure 4.7. The optimal bias is -189.8 V.

(Figure 4.1). It was assumed that the LED resistance was constant. However, an LED is a diode and the resistance of a diode varies with applied voltage (recall Figure 2.7, the current-voltage curve of an APD). The LED intensity was inversely proportional to the resistance of the *entire* circuit. In this chapter, a better definition of the intensity is the current through the diode. However, the current was not measured for this experiment. The plot of the APD response as a function of $1/R_V$ was nonlinear in Figures 4.5 and 4.6. A log (S/N) *vs.* log (Intensity) at the various APD biases is plotted in Figure 4.7, where the signal and noise was in Volts and Intensity was $1/R_V$.

The optimal APD bias was determined from the linear approximation analysis in Figure 4.7. The slopes and intercepts at a given APD bias are given in Figure 4.7. The best signal-to-noise ratio at low light levels occurred at a bias of -189.8 V. The signal-to-noise ratios were also calculated at a given APD bias and light intensity. The signal-to-noise ratios are plotted as a function of APD bias in Figure 4.8. A maximum signal-to-noise ratio appeared at the APD bias -189.8 V.

Figure 4.7 also demonstrated the light range for the APD at a given bias. At the low APD biases, -179.4 to -187.8 V, the APD had difficulty detecting signals at log(Intensity) values less than -1.8 (Intensity = 0.015 $k\Omega^{-1}$). At high APD biases, -188.8 V to -191.0 V, the APD saturated the oscilloscope input at log (Intensity) values greater than -1.5 (Intensity = 0.03 $k\Omega^{-1}$).

This entire section was devoted to determining the optimal APD biasing voltage. According to EG&G, the optimum signal-to-noise occurs at a gain M where the total detector noise equal the input noise of the external amplifier [3, 5]. However, without knowing M and the responsivity $R(\lambda)$, the optimal bias of APDs had to be determined experimentally.

Based on this signal-to-noise study, the bias was determined to be -189.8 V for this APD with an external gain of 10^{10} V/A. At this bias, the APD was able to detect the low light levels but not at the expense of increased detector noise. It will also be shown in the following section that this APD was signal-shot-noise limited at a bias of -189.8 V.

All of the APDs on the 16-capillary instrument were biased in a similar manner. The signal-to-noise ratios were measured as function of APD bias at a constant light

intensity. The bias with the maximum signal-to-noise ratio was chosen as the optimal setting. The optimized APD biases are listed in Section 3.3.2.4.

4.3.1.3. Characterization of the Noise in APDs

The noise in the APD was characterized with log (S/N) *vs.* log (S) plots. Figure 4.9 shows the log (S/N) *vs.* log (S) plots at the various APD biases with external gain of 10^9 V/A. At a bias of -179.4 V, the slope of the line was 0.93, indicating that the noise was constant and independent of signal. As the bias increased, the slope of the line decreased until it reached a value of 0.5 at APD biases of -191.0, -191.4 and -191.7 V. A slope of 0.5 in a plot of log (S/N) *vs.* log (S), indicated a square root relationship between S/N and S, which is a signal-shot-noise-limited situation. At a first glance, the optimal APD bias may appear to be -191.4 and -191.7 V, because of the shot-noise limit. However, we know that the S/N levels decreased with increasing bias.

Figure 4.10 shows the log (S/N) *vs.* log (S) plots at various APD biases with external gain of 10^{10} V/A. Increasing the bias decreased the slope of 0.68 at APD bias of -179.4 V to a slope of 0.42 at bias of -191.0 V. The change in the slopes of the lines was less dramatic compared to the slopes at a gain of 10^9 V/A. I am uncertain why the slope of the line jumped to 0.68 at APD bias -188.8V. At -189.8 V, the chosen optimum bias for APD #8 in the previous section, the slope of the plot was 0.5, indicating the APD was signal-shot-noise-limited.

Figure 4.11 shows the log (S/N) *vs.* log (S) plot of the SPCM. Interestingly enough, the slope of the line was not 0.5 but 0.75, indicating that the SPCM was not signal-shot-noise-limited. The limiting noise may arise from the frequency-to-voltage converter noise used with this device.

4.3.1.4. Calculation of Noise in APD

A comparison was made between the noise expected in the APDs and the experimental noise.

The manufacturer's specifications of a typical APD at room temperature are found in Table 4.1. At 22°C, the APDs have a dark current of 30 nA. At a 1-Hz bandwidth, the

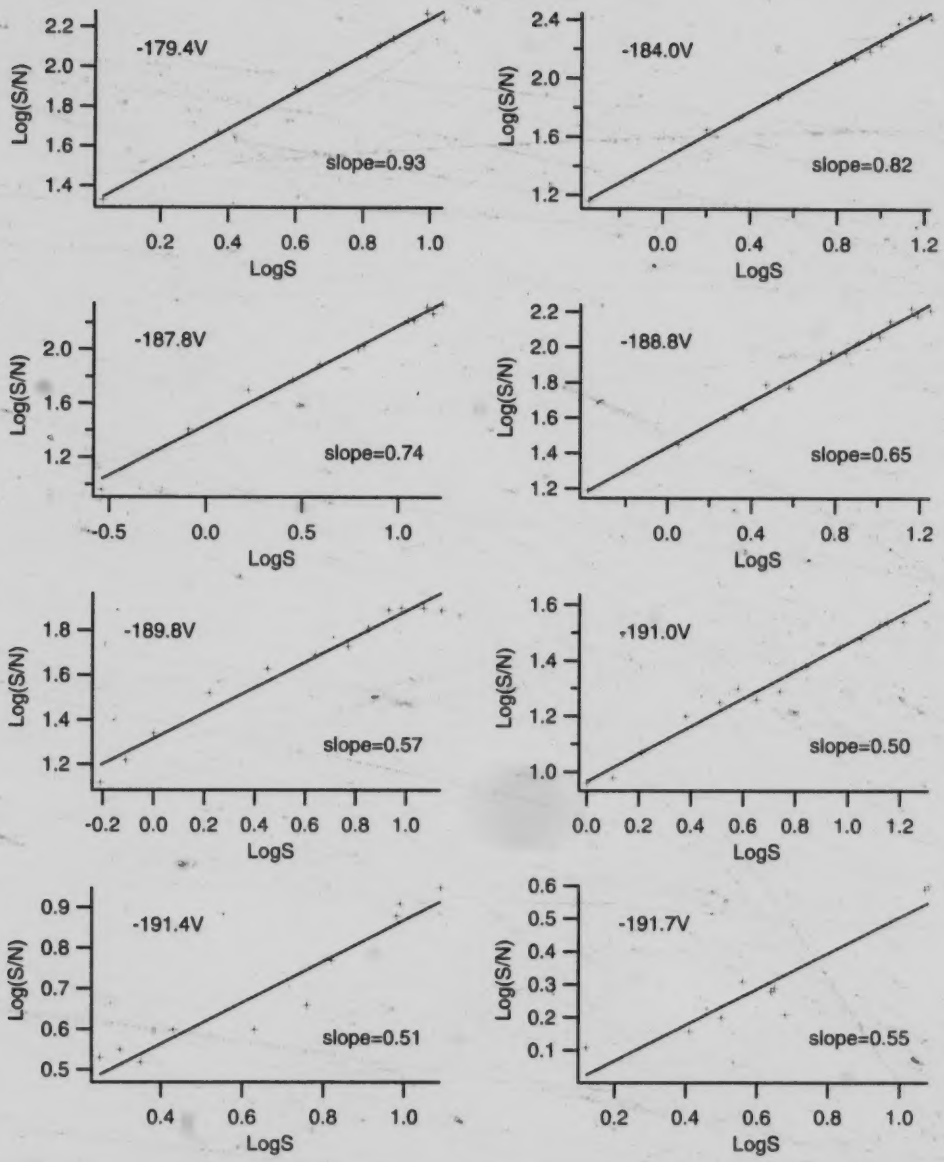


Figure 4.9. Log (S/N) *vs.* Log (S) at External Gain of 10^9V/A

The data from Figure 4.5 are plotted as a log signal-to-noise *vs.* log signal with increasing APD bias. At lower biases, the noise is constant and independent of signal. At higher biases, the signal-to-noise ratio decreases but the APD is signal-shot-noise limited.

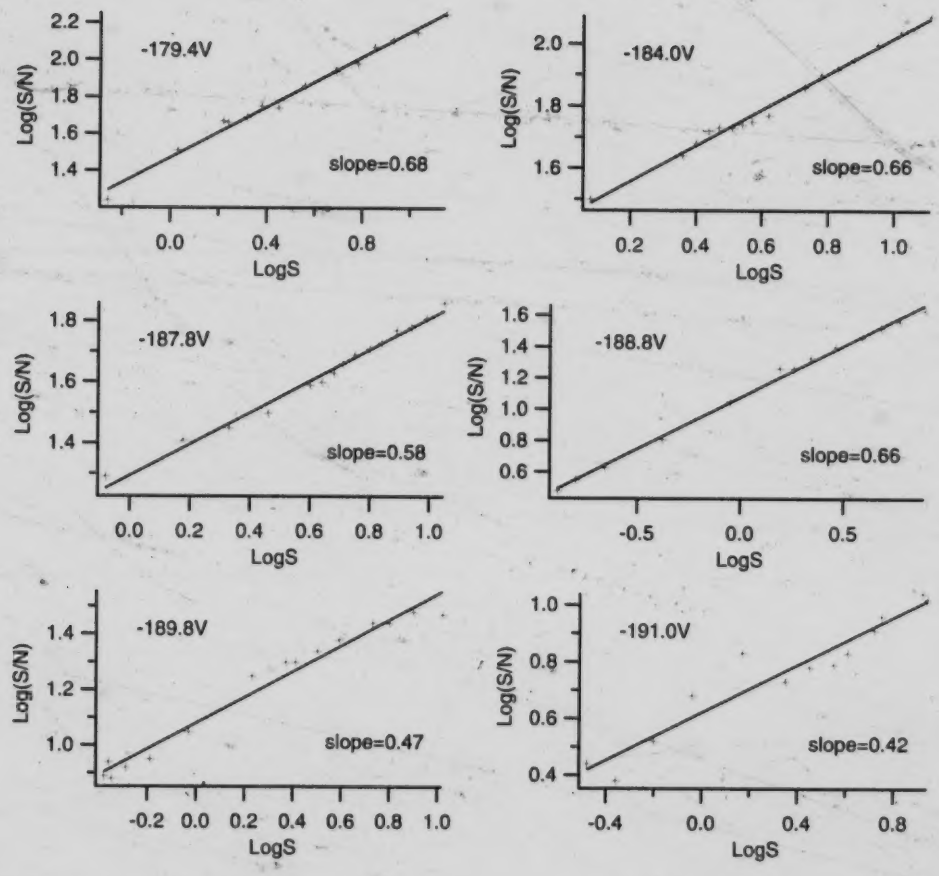


Figure 4.10. Log (S/N) *vs.* Log (S) at External Gain of 10^{10}V/A

The data from Figure 4.6 are plotted as a log signal-to-noise *vs.* log signal with increasing APD bias. At lower biases, the noise is constant and independent of signal. At higher biases, the signal-to-noise ratio decreases but the APD is signal-shot-noise limited.

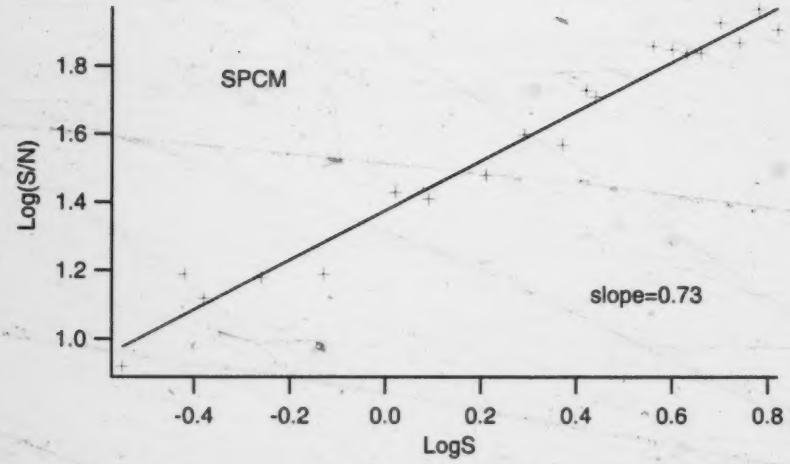


Figure 4.11. Log S/N *vs.* Log S of Single Photon Counting Module.

dark current noise is 1.1×10^{-13} A, which is equivalent to an incident photon flux of 5500 photons/s.

Table 4.1. Specifications of EG&G SiAPD at 22°C [8]

	Typical EG&G SiAPD (C30902SQC)
Photosensitive Surface Diameter	50 μm
Breakdown Voltage, V_{BR}	-225 V
Temperature Coefficient of V_R	0.7 V/°C
Linear Mode Gain	250
Quantum Efficiency at 550 nm	50%
Responsivity at 550 nm, Gain=250	55 A/W
Dark Current	30 nA
Spectral Noise Current Density	1.1×10^{-13} A/Hz$^{1/2}$

The values in Table 4.1 were used to calculated the values at -20°C, in Table 4.2. At -20°C, the dark current is calculated as 2 nA because the dark current roughly halves with every 10°C decrease. The excess noise factor is seven at a gain of 250 and an ionization value of k = 0.02 for silicon APDs [3]. The noise in the APD, $\sigma_{Dark,Shot,}$ is ~3×10^{-14} A at 1 Hz bandwidth, estimated from Equation 4.10. At 550 nm, the noise is equivalent to an incident photon flux of ~1300 photons/s at the detector.

Table 4.2. Calculated Specifications of SiAPD at -20°C

	SiAPD
Breakdown Voltage, V_{BR}	-197 V
Linear Mode Gain	250
Quantum Efficiency at 550 nm	50%
Dark Current	2 nA
Spectral Noise Current Density	3×10^{-14} A/Hz$^{1/2}$

The noise of the first and second generation APD circuits is calculated and listed in Table 4.3. The noise current in the APD is 7.9×10^{-14} A and 8.5×10^{-14} A at 7 Hz and 8 Hz bandwidths respectively using the spectral noise current density at -20°C. The noise in the APD is calculated by multiplying the noise current by the gain of the amplifiers and

gain in the low-pass filters. The noise in the amplifier σ_{ar} is calculated from Equation 4.2. The total detector noise σ_D is calculated from Equation 4.6.

Table 4.3. Expected APD Detector Noise at -20°C

	1st Generation APD	2nd Generation APD
Amplifier Gain	10^9 V/A	10^{10} V/A
Amplifier Feedback Resistor	10^9 Ω	10^{10} Ω
Low-Pass Filter Gain	10	20
Bandwidth, Δf	7 Hz	8 Hz
APD Noise, $\sigma_{Dark,Shot}$	0.8 mV	17 mV
Amplifier Noise, σ_{ar}	0.01 mV	0.04 mV
Total Detector Noise, σ_D	0.8 mV	17 mV

In the first generation instrument, the detector noise was a few mV (Section 2.3.2.3) and in the second generation instrument, the total detector noise was ~ 30 mV (Section 3.3.2.3). These experimentally determined values agree with the expected total detector noise values.

4.3.2. Limit of Detection

4.3.2.1. Fiber Optic Core Comparison on 5-Capillary Instrument

The 5-capillary DNA sequencer with a sheath-flow cuvette and laser induced fluorescence detection was described by Zhang [1]. The 5-capillary instrument had a detection limit of 130 molecules, a value that could not be attained on the 16-capillary instrument. A direct comparison of the limit of detection values between the two instruments was not very informative. It was difficult to pinpoint if less efficient optics or poorer detectors caused the sensitivity difference in the instruments. The differences in the optics and the detectors had to be distinguished to determine the factors limiting the performance of the 16-capillary instrument.

The optics, specifically the GRIN/fiber optic connections, from the 5-capillary and 16-capillary instruments were compared. The fluorescein dye was injected on capillaries in the 5-capillary instrument. The fluorescence was imaged onto GRIN lenses connected to two different fiber optics from the two instruments. The GRIN lens from the 5-capillary instrument was connected to a 100 μm diameter fiber optic core. The GRIN

lens from the 16-capillary instrument was connected to a 50 μm diameter fiber optic core. It was uncertain if the difference between the two core diameters influenced the performance of the detectors.

The signal-to-noise ratios for the two different fiber cores are listed in Table 4.4, where the signal-to-noise was the height of the fluorescent peak divided by the noise in the background. The calculated LODs (S/N=3) with the same detector and two different GRIN-fiber optics are listed in Table 4.5. The data were smoothed using Savitzsky-Golay second order filter of 21 points.

Table 4.4. Signal-to-Noise Ratios as Function of Fiber Optic

	Raw Data		Smoothed Data	
[fluorescein] (M)	1×10^{-9}	2×10^{-10}	1×10^{-9}	2×10^{-10}
S/N of 100 μm fiber optic	400	150	630	300
S/N of 50 μm fiber optic	220	70	400	140
$\frac{(S/N)_{100\mu m}}{(S/N)_{50\mu m}}$	1.8	2.1	1.6	2.1

Table 4.5. LOD Values of Two Fiber Optics

	raw data	smoothed
[fluorescein] (M)	2×10^{-10}	2×10^{-10}
LOD of 100 μm core (molecules)	8000	4000
LOD of 50 μm core (molecules)	16000	8000

The 100 μm core had an area that is four times larger than the 50 μm core. There was in Table 4.4, however, only a twofold increase in S/N ratio at a given fluorescein concentration (LOD is inversely proportional to S/N). This twofold increase possibly represented the signal-shot-noise-limit of the SPCM detectors, where the noise was proportional to the square root of the signal. Thus a fourfold increase in signal resulted in an increase of two in the noise. While the SPCM was shown not to be signal-shot-noise-

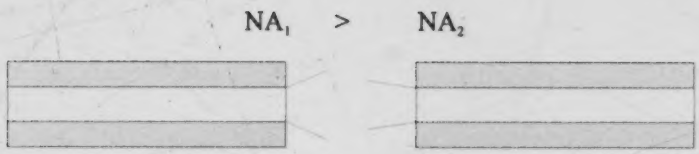
limited in Section 4.3.1.3, it may be in this experiment on the 5-capillary instrument. A different frequency-to-voltage converter built by Zhang was used in this experiment [9].

The volume (Vol_{inj}) of fluorescein injected onto the capillary was calculated to be ~3 nL from Equation 4.15. The LODs calculated in this experiment were 30 times higher than the values calculated by Zhang in his thesis. Zhang reported 130 ± 30 molecules and 2×10^{-13} M fluorescein detection limit for the 5-capillary instrument. The main difference in the LOD value can be attributed to the use of the peak height to calculate the signal-to-noise ratio. In poorer separations, broader peaks had lower plate numbers (N) and lower peak heights. I obtained a theoretical plate number of 50000. Zhang obtained an average N = 330000 [1]. This lower plate number was attributed to significant band broadening from the injection of high concentration fluorescein solutions. This reason may also explain why the signal-to-noise ratios in Table 4.4. were not proportional to fluorescein concentration. Obviously the separation was not ideal but this experiment served to demonstrate the collection efficiency difference between the two fiber optic sizes.

It was assumed that the connections between the GRIN-fiber optic to APD-fiber optic and GRIN-fiber optic to SPCM were similar. There may however be an insertion loss or a loss of optical power through a connector as shown in Figure 4.12 [10]. The SPCM has a 200 μm diameter circular photosensitive area with a peak photon detection efficiency greater than 60% at 633 nm [11]. It was assumed that the 50 μm fiber optic was well aligned and the core overlapped with the center of the photosensitive area. Poor alignment would result in the SPCM receiving less light.

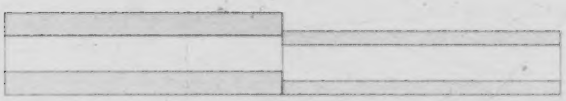
4.3.2.2. LOD Measurements on 16-Capillary Instrument

The limit of detection from the injection of 2×10^{-11}M fluorescein was 15000 molecules (8 pM, 21% rsd, n=5) from the raw data and 10000 molecules (5 pM, 31% rsd, n=5) from the smoothed data. Figure 4.13 shows the smoothed peaks used to calculated the LOD and their calculated plates, N, which average to about 200000 (or plate height, H, of 2 μm for a 42.5 cm long capillary).



a. Numerical Aperture Mismatch

b. Core Diameter Mismatch



c. Cladding Diameter Mismatch

Concentricity Eccentricity Ellipticity

d.

Figure 4.12. Intrinsic Factors of Fiber Optic Connector Loss.

Light power is lost as a signal is sent through a connector. Intrinsic losses occur due to defects in the fiber optics. In a), light is lost when the numerical aperture is different for the two fibers. The numerical aperture indicates the maximum light input angle of a fiber. Light at an angle larger than the maximum input angle will not be transmitted. This loss occurs from when light is transmitted from a high NA to lower NA fiber-optics. In b), core diameter mismatch results in loss of light power when light is connected through two fibers with different diameters. In c), if the cladding diameter is different, the cores of the fibers will not be aligned. In d), the radial position of the core in the cladding is important. Ideally, the fiber core and cladding are concentric. However, if the core is eccentric, it will not match with a concentric fiber optic. Also if the cores are elliptic and do not align, light will be lost.

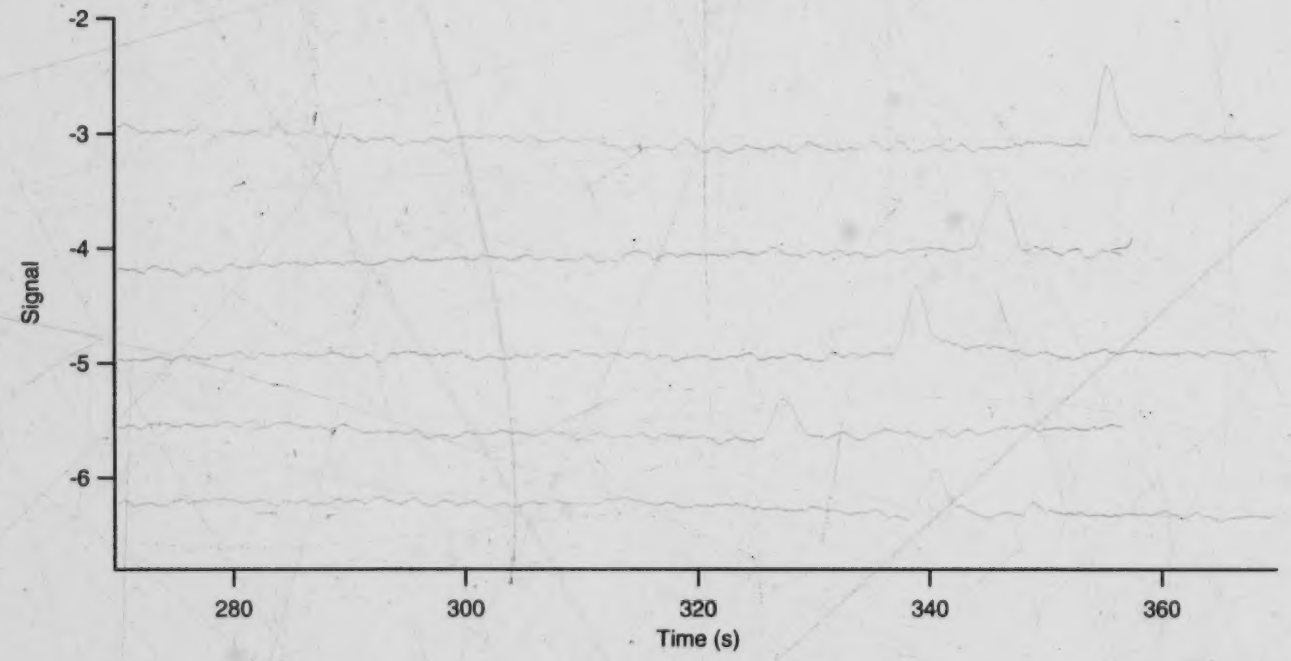


Figure 4.13. Injection of 2 x 10^{-11}M Fluorescein.
Five replicate injections of 39 000 molecules of fluorescein. Smoothed data.

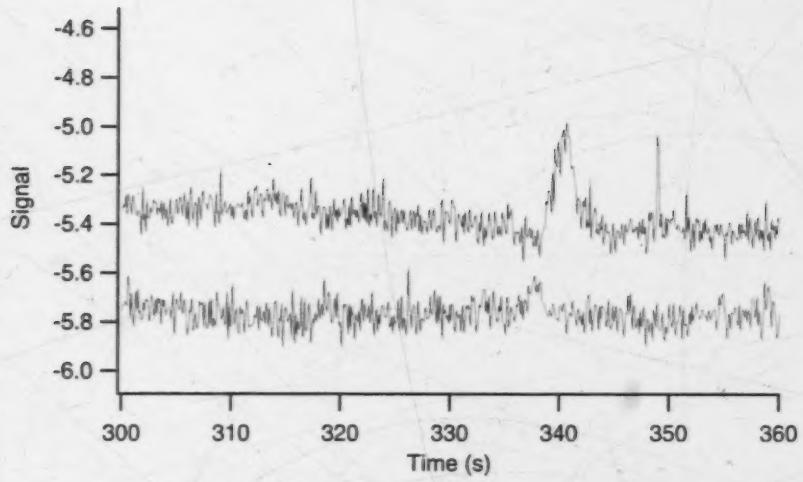


Figure 4.14. Raw Data of the Injection of 2 x 10^{-11}M Fluorescein.
Top: Injection of 39 000 molecules of fluorescein. Bottom: Siphon peak.

Figure 4.14 shows an unprocessed peak from the injection of 2×10^{-11} M fluorescein and a blank peak illustrating the extent of siphoning. Siphoning of the sample in the capillary occurs because of the effect of gravity on height differences between the sample level and the buffer level in the waste container. If the sample level is higher than the waste level, the sample is passively loaded onto the capillary. Reverse siphoning, where the sample is pushed back out of the capillary, may occur if the sample level is lower than the waste level. To calculate the detection limits, the exact amount of sample loaded electrokinetically and passively onto the capillary must be known. In this experiment, the buffer level in the waste container was set slightly lower than the sample injection height to ensure a small amount of siphoning in the capillary. The height of the siphon peak was then subtracted from the height of the height of the injection peak.

The 16-capillary concentration detection limit was 25 times poorer than the 5-capillary instrument. From Section 4.3.1.2, the signal-to-noise ratios of the optimized APDs were similar to the SPCM; the detectors could not be the limiting factor. Any differences in sensitivity between the 5-capillary instrument and the 16-capillary instrument must be the result of the collection efficiency of the optics. The 50 μm fiber optic core accounted for a two-fold difference in LOD. Other differences in the values may be from the lower collection efficiency of the two microscope objectives. The collection efficiency was 5.3% for the 0.45 NA microscope objectives and 6.7% for the 0.5 NA microscope objective. In all the16-capillary APD detectors received 40% of the light that the 5-capillary detectors did.

4.4. CONCLUSIONS

It was demonstrated that the APD sensitivity increased with increasing APD bias. However the signal-to-noise ratio decreased with increasing APD bias. To improve the signal-to-noise ratio, the external gain of the APDs was increased. At this increased external gain, the second generation APDs had similar signal-to-noise ratios as the SPCM. The APDs were also shown to be signal-shot-noise-limited.

The performance of the APDs were similar to the SPCMs. Thus any differences between the APD and SPCM were determined to be a result of differences in the optics of the 16-capillary and 5-capillary instruments.

In this chapter, I reported 5×10^{-12} M fluorescein detection limit for the 16-capillary instrument. In the next chapter, I use the improved instrument for sequencing the heat shock protein gene in several *Staphylococcus* species.

4.5. REFERENCES

(1) Zhang, J. Z. Z. Ph.D., University of Alberta, Edmonton, **1994**.

(2) Ingle Jr., J. D.; Crouch, S. R. *Spectrochemical Analysis*; Prentice-Hall, Inc.: Toronto, **1988**.

(3) Avalanche Photodiodes: A User's Guide, EG&G Optoelectronics, Canada.

(4) Lee, T. P. In *Fiber Optics*; Daly, J. C., Ed.; CRC Press, Inc.: Boca Raton, Florida, **1984**, pp 123-150.

(5) MacGregor, A. *Photonics Spectra* **1991,** *February*, 139-146.

(6) Boumans, P. W. J. M. *Analytical Chemistry* **1994,** *66*, 459-467A.

(7) Huang, X.; Gordon, M. J.; Zare, R. N. *Analytical Chemistry* **1988,** *60*, 375-377.

(8) Silicon Avalanche Photodiodes C30902E, C30902S, C30921E, C30921S Data Sheet, RCA ElectroOptics.

(9) Zhang, J. Z. Z., Frequency-to-Voltage Converter, **1994**.

(10) Ajemmian, R. G. *Optics and Photonics News* **1995,** *6*, 32-36.

(11) Single Photon Counting Module SPCM-AQ Series Provisional Data Sheet, EG&G Optoelectronics, Canada, **1994**.

Chapter 5: *Staphylococcus* HSP60 Gene Sequencing

5.1. INTRODUCTION

Recent advances in DNA sequencing chemistry and technology make it possible to sequence large amounts of DNA in short periods. Multiple-capillary electrophoresis technology will play an important role in the generation of vast amounts of DNA sequence because the instrumentation is amenable to automation. This chapter demonstrates the potential of multiple-capillary DNA sequencing and presents DNA sequence data for a proposed genotype diagnostic to identify bacterial species.

This microbial diagnostic relies on differences in the DNA sequences in the ubiquitous and highly conserved Heat Shock Protein-60kDa gene (HSP60) of different *Staphylococcus* species. The HSP60 gene has both conserved and variable regions. A set of PCR primers was designed to bind to the conserved regions flanking a variable region of the HSP60 gene. Amplification of the variable region by Polymerase Chain Reaction (PCR) resulted in a 600 bp product. The same primer set could amplify 600 bp PCR products in different *Staphylococcus* species. Sequencing the PCR products confirmed the sequence variations in the gene for species identification. The sequences will be used to create a library of HSP60 gene sequences of different *Staphylococcus* species for microbial identification.

5.1.1. Phylogenetic Tree Construction

How is a bacterial species identified knowing only a gene sequence? After sequencing the gene, the next step would be to run the DNA sequence (or translated peptide sequence) through a homology search program. Matches indicate that the DNA is similar to a sequence in the data banks. Identification of the bacterial species relies on the presence of the target gene sequence in a data bank. Identification of a novel bacterium may be based on its similarity and homology to related organisms, which is presented as a phylogenetic tree. The tree would also infer evolutionary histories of the bacterial species.

Phylogenetic trees, or dendrograms, depict an evolutionary relationship among organisms. The relationship can be phenetic or cladistic. Phenetic relationships are

similarities based on phenotypic characteristics and cladistic relationships contain information about ancestry and are used to study evolutionary pathways [1]. Another note on terminology: similarity is an empirical observation about the sequences but does not imply any evolutionary relationship. Homology implies that there is a common ancestry between the organisms [2]. A phylogenetic tree is composed of branches, also called nodes and illustrated in Figure 5.1. If the ancestor from the earliest point in time is unknown, the tree is "unrooted".

A few words need to be said about the types and rates of mutations. There are four basic types of mutations at the DNA level. They are: *substitutions* of a nucleotide for another nucleotide, *deletion* of nucleotides, *insertion* of nucleotides and *inversion* of a nucleotide, where the base is changed to its complement. The substitution mutations can be broken down into: *transition*, the substitution of a purine for purine (A $\leftrightarrow$ G) or pyrimidine for a pyrimidine (T $\leftrightarrow$ C) and *transversion*, the substitution of a purine for pyrimidine or *vice versa*. In the coding regions of genes, the majority of mutations are substitutions; transition substitutions occur more frequently than transversions. It has been observed that substitution mutations occur more in noncoding regions compared to coding regions and the mutation rates among all four nucleotides are not equal [3]. In addition, the third codon position in a gene is more frequently substituted than the first or second codon positions [4].

How is a phylogenetic tree constructed? A sequence by itself has no meaning. First the DNA or peptide sequences are aligned for similarity and homology. Second, the similarity between sequences is reduced to a "distance" relating all pairs of sequences. Third, a tree is constructed based on paired distances. Fourth, the tree stability can be evaluated by "bootstrapping".

Sequence Alignment

There are several methods to compare the similarity of different sequences: dot plots, global alignments, and local alignments. Most of the algorithms applied to DNA sequence alignments were originally derived from methods used to analyze amino acids. Various algorithms for nucleotide sequence alignments were reviewed by Davison [5].

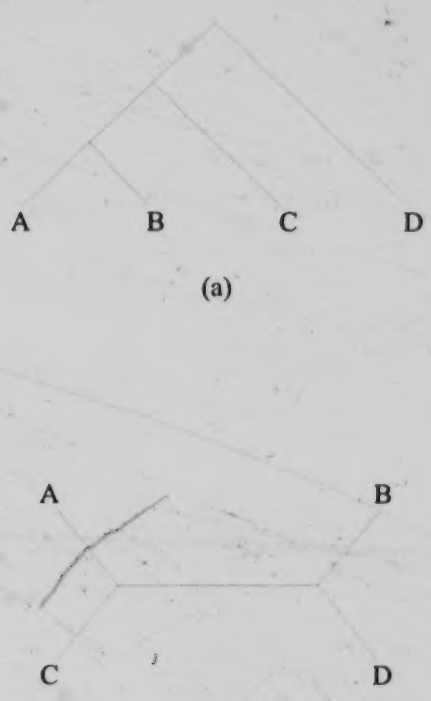


Figure 5.1. Phylogenetic Trees.
Phylogenetic relationships are usually presented in the form of a tree. The tree is composed of branches where the units diverge. The lengths of the branches indicate the distance between the units, the time since two units diverged. In (a), the tree is rooted. A rooted tree conveys a temporal relationship between the organisms, in relation to a common ancestor. In (b), the tree is unrooted, merely reflecting the distances between units without indicating which is the common ancestor.

Global alignments match two homologous regions, taking into account total lengths of the sequences. Gaps are used to compensate for inserted or deleted bases. To avoid aligning sequences that are dissimilar, gaps are penalized. The number of gaps are penalized more heavily than the size of gap based on the assumption that insertion and deletion events occur on individual nucleotides rather than short sequences. This alignment was developed for peptide sequences to maximize the degree of global similarity [6] and then for DNA sequences [7].

Global alignments can miss regions of subsequence similarity. Local alignments find areas of local sequence similarity and then build longer subsequence alignments for more distantly related sequences. Smith described an algorithm to find a pair of segments with the greatest similarity from two long sequences [8].

Basic Alignment Search Tool (BLAST) is a commonly used local alignment program [9] and a variation is available from the National Center for Biotechnology Information (GenBank, http://www.ncbi.nlm.nih.gov). The NCBI BLAST program looks for a High-Scoring Pair (HSP), which consists of aligning short, seed sequences of equal length and that have local sequence similarity. The alignment score is calculated from a mutation probability matrix and considers differential mutation rates [10, 11]. The matrix is composed of points for a pair of matching residues, and penalties for mismatching residue. Gaps are not allowed. The alignment score must be greater than or equal to some threshold value to form a seed sequence. Once a seed HSP pair is found, longer HSPs are sought, extending in both directions from the seed sequence. The HSP extends until the cumulative alignment score falls below a threshold value, the score equals zero, a negative-scoring residue is aligned, or a sequence's end is reached.

The HSPs are evaluated to see if they are statistically significant to determine if the similarity could have arisen by chance. The theory of Karlin and Altschul gives a method for constructing such a measure [12, 13]. The expected frequency of a chance occurrence is:

$$E = KNe^{-\lambda S} \qquad (5.1)$$

where E is the expected frequency of a chance occurrence of an HSP having an alignment score S, K and λ are Karlin-Altschul parameters and $N = m \cdot n$, where m and n are the

lengths of the paired sequences. The probability of a chance occurrence, P, is expressed as:

$$P = 1 - e^{-E} \tag{5.2}$$

A statistically significant alignment occurs when E approaches 0 and P approaches 0.

Distances

The similarity of two aligned sequences is commonly calculated as:

$$\text{Similarity} = \frac{M}{L} \tag{5.3}$$

$$L = M + U + w_G G \tag{5.4}$$

where L is the effective length of the aligned sequences, M is the number of positions containing identical residues, U is the number of number of positions containing different residues, and G is the number of gaps in the aligned sequences. w_G is the weight given to gaps and can vary from zero (gaps are ignored) to one (gaps are equivalent to a mutation). Similarity is expressed as a value from 1 to 0 (100% to 0%). For example, one sequence is 93% identical to a sequence from a homologous species [4]. The distance between two organisms may be derived from the similarity.

The distance calculation is more ambiguous and depends on the mutation rate algorithms. Generally, identical objects are separated by zero distance, but there is no consensus on the maximum value of distance; it may be infinite [4]. An example of an evolutionary distance (the number of substitutions per sequence position) calculation can be represented by the following formulae [2]:

$$D = 1 - \text{Similarity} \tag{5.5}$$

$$d = -b \ln\left(1 - \frac{D}{b}\right) \tag{5.6}$$

D is the sequence dissimilarity, d is the estimated evolutionary distance between a pair of sequences and b is a value related to the frequency of substitution depending the mutation model used. In the Jukes and Cantor model, all four nucleotides are equally likely to substitute [14]. The Kimura Two-Parameter model considers unequal rates of transitions and transversions to calculate evolutionary distance [15].

Phylogenetic Tree Building

Distance calculations and tree building are intertwined analyses. There are several methods used to construct phylogenetic trees: distance-matrix, parsimony, and maximum likelihood.

The distance-matrix method to build trees is based on pair-wise distances. The distances usually refer to number of changes between the pair. The most commonly used distance-matrix method is Neighborhood-Joining, which assumes the same rate of evolution applies to each tree branch [16]. Basically the method begins with a star-shaped phylogeny and neighbors are pairs of species that form a tree of the shortest length. The distances between the neighbors is additive and equals the sum of the lengths of the branches. This method is fast, simple, and allows the use of statistical tests such as bootstraps. It is however somewhat imprecise because it leads to loss of information by reducing the data to a single number, the distance.

The parsimony method [3] does not look at the distances between pairs and as a result does not require the use of evolutionary models of frequency of substitutions. The method evaluates all the possible trees and gives each tree a score. The score is related to the number of evolutionary changes required to create the given tree. The most parsimonious tree is the one wherein the organisms evolved with the fewest evolutionary transformations. This method works well if the substitution of nucleotides is constant or small between the branches. Problems in tree construction can occur when only few uninformative mutations occur, such as in highly conserved sequences.

Maximum likelihood uses all the data instead of reducing data to distances. Unlike parsimony, it uses a statistical model of evolution [17]. It also considers that changes are more likely to occur in a long branch than a short one. First a tree is formed and then the branch lengths are chosen to maximize the likelihood of that tree. The probabilities are compared between different trees. The biggest disadvantage of this method is the computational effort required to form all the possible trees. The number of unrooted trees that are bifurcating (two branches joining at an interior branch) for n species is:

$$\text{Number of trees} = \frac{(2n-5)!}{\left(2^{n-3}(n-3)!\right)} \tag{5.7}$$

For ten species, there are 2 027 025 possible trees!

Bootstrapping

The bootstrap technique puts an empirical confidence level on the constructed tree. Basically, the data is resampled by successively dropping individual data points or by replacement of the data from the original set [18]. The "bootstrapped" data is realigned and new phylogenetic trees are constructed. The bootstrap analysis determines the stability of the data to changes in data set. The overall topology of a robust tree construction is not likely to change with the removal of data from the set. The "majority rule" consensus tree consists of the grouping of species that appear in the majority of the bootstrapped trees. The bootstrapping technique is commonly used for evaluating the constructed tree [19-22].

Phylogenetic Identification

An example of phylogenetic bacterial identification was reported in the newly proposed *Staphylococcus* species, *S. lutrae* [21]. Biochemical tests, such as enzyme-activity and antibiotic susceptibility, pointed towards a coagulase-positive *Staphylococcus*-like species. Comparative 16S rRNA gene sequence analysis established its relation to other *Staphylococcus* species. There are other examples of phylogenetic characterizations confirming the identity of questionable species [23] and even the reclassification of bacterial species [20].

A word of caution: the phylogenetic trees constructed from the sequence of a gene are in reality "gene trees" and not species trees [3]. Trees constructed from different genes of the same species may not be the same. Polymorphisms can exist within a species and the genes may have not have the same phylogenetic history. Instead of relying on a single genotype to classify species, there is a growing movement toward "polyphasic" classification. This method promotes the identification of bacteria by

considering both phenotypic, genotypic and phlyogenetic information [24] and the development of the appropriate software tools and databases [25].

5.2. EXPERIMENTAL

5.2.1. Bacterial Isolates

Six reference strains of *Staphylococcus* species were obtained from Dr. Anthony Chow, Department of Medicine, Division of Infectious Diseases, University of British Columbia. The strains of *Staphylococcus* were: *S. aureus* 8325-4, *S. epidermidis* 9759, *S. haemolyticus* ATCC 29970, *S. lugdunensis* CRSN 850412, *S. saprophyticus* KL122, and *S. schleiferi* ATCC 43808. They were taxonomically classified as: Eubacteria; Firmicutes; Low G+C gram-positive bacteria; Bacillaceae; Staphylococcus.

5.2.2. HSP60 PCR and Cloning

A schematic of HSP60 PCR and cloning is shown in Figure 5.2. Briefly, a single *Staphylococcus* colony was selected and a portion of the HSP60 gene was amplified by PCR. The HSP60 PCR product was ligated into a plasmid vector, transformed and grown in *E. coli*. The positive colonies, antibiotic resistance and white, were screened for the presence of the HSP60 insert by PCR. If the insert was present, the screening PCR product was used as a template for DNA sequencing.

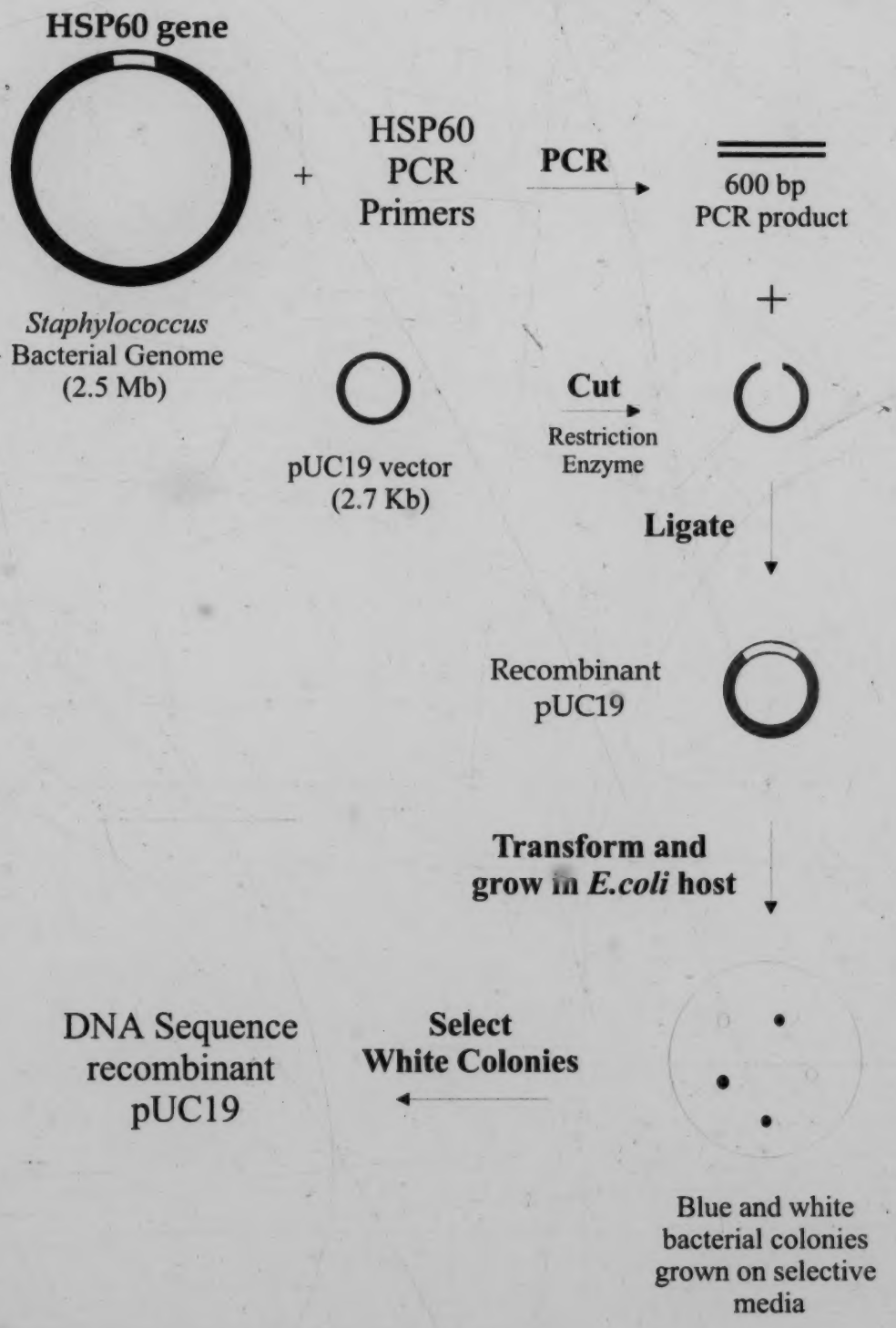
5.2.2.1. Boiling Lysis HSP60 PCR

The bacteria were plated and grown overnight at 37°C on Brain Heart Infusion (BHI, Becton Dickinson, Cockeysville, MD) agar plates. The DNA was prepared by the "boiling/lysis" PCR preparation. A single colony was touched with a sterile P10 pipet tip. The tip was dipped into 20 µL autoclaved water and the tube was boiled for 2 min. The cellular debris were pelleted with a quick spin in the microcentrifuge and 10 µL of the supernatant was used for HSP60 PCR reaction.

The PCR was a modified version of the original PCR [26]. The 5' and 3' PCR primers used were called H279, with the sequence, 5'-GAATTCGAIIIIGCIGGIGA(TC)GGIACIACIAC-3', and H280, with the sequence, 5'-CGCGGGATCC(TC)(TG)I(TC)(TG)ITCICC(AG)AAICCIGGIGC(TC)TT-3'. 42 pmol

Figure 5.2. HSP60 PCR and cloning

The bacterial DNA was lysed from the cell and the HSP60 gene was amplified by PCR. The resulting 600 bp PCR product was ligated into plasmid pUC19. The recombinant plasmid, pUC19 with HSP60 PCR product, was transformed into an *E. coli* host. The pUC19 plasmid conferred antibiotic resistance to the bacteria, allowing only bacteria containing pUC19 to grow on selective media. Bacteria containing recombinant pUC19 were white and bacteria containing pUC19 without the insert were blue. The white colonies were selected, and the DNA inserted into pUC19 was sequenced.


HSP60 gene
+
HSP60
PCR
Primers
PCR
600 bp
PCR product
+
Staphylococcus
Bacterial Genome
(2.5 Mb)
pUC19 vector
(2.7 Kb)
Cut
Restriction
Enzyme
Ligate
Recombinant
pUC19
Transform and
grow in *E.coli* host
DNA Sequence
recombinant
pUC19
Select
White Colonies
Blue and white
bacterial colonies
grown on selective
media

of H279 and 42 pmol of H280 of were added to 10 µL of genomic bacterial DNA, 20 nmoles dNTPs, 150 nmoles $MgCl_2$, 10 µl of 10×PCR reaction buffer and 5 U of Taq DNA Polymerase (Gibco BRL, Gaithersburg, MD). The final volume of reaction was 100 µL. A negative control was prepared with all of the above reagents except the bacterial DNA. The thermal cycling conditions were: 3 min at 95°C, 40 cycles of 30 s at 94°C, 30 s at 37°C and 1 min at 72°C, and last cycle of 5 min at 72°C (PTC-100 Programmable Thermal Controller, MJ Research, Watertown, MA).

5 µL of the PCR reaction was run on 0.8% agarose gel in 1×TAE buffer (40 mM tris, 20 mM acetic acid and 1.0 mM pH=8.0 EDTA) with ΦX174 RF DNA/*Hae*III size ladder (Gibco BRL) at 100 V for 1 h. The DNA was visualized with ethidium bromide (EtBr) under UV light for the 600 bp products as shown in Figure 5.3.

5.2.2.2. Cloning the HSP60 PCR Products into pUC19

The H279 primer had an *Eco*RI restriction enzyme digest site, indicated by the underscoring in the above sequence. H280 primer had a *Bam*HI digest site, also indicated by underscoring. 40 µL of PCR product was double-digested with 10 Units each of *Eco*RI and *Bam*HI (Gibco BRL). The total reaction volume was 50 µL. 1 µg of plasmid vector pUC19 was double-digested with *Eco*RI and *Bam*HI. The digests were incubated at 37°C for 1 h. The digested products were cleaned by QIAquick PCR Purification Kit, (Qiagen Inc, Chatsworth, CA) resuspended in 10 mM Tris-HCl (pH=8.0) and DNA concentration was measured with a fluorimeter (TD760, Turner Designs, Sunnyvale, CA) and EtBr. Concentrations ranged from 10 to 50 ng/µL.

The ligation reaction was prepared with 80 ng of *Eco*RI/*Bam*HI digested PCR, 20 ng of *Eco*RI/*Bam*HI digested pUC19 (approximately 20:1 ratio of PCR product to pUC19) and 1 U of T4 DNA ligase enzyme (Gibco BRL). The total reaction volume was 15 µl. A negative control of pUC19 without PCR product was also prepared. The ligation reactions were incubated at 16°C overnight in a cooled incubator.

The recombinant pUC19 vector was transformed into *E. coli* strain DH5α (United States Biochemical, USB, Cleveland, OH). 7.5 µL of the ligation reaction was added to

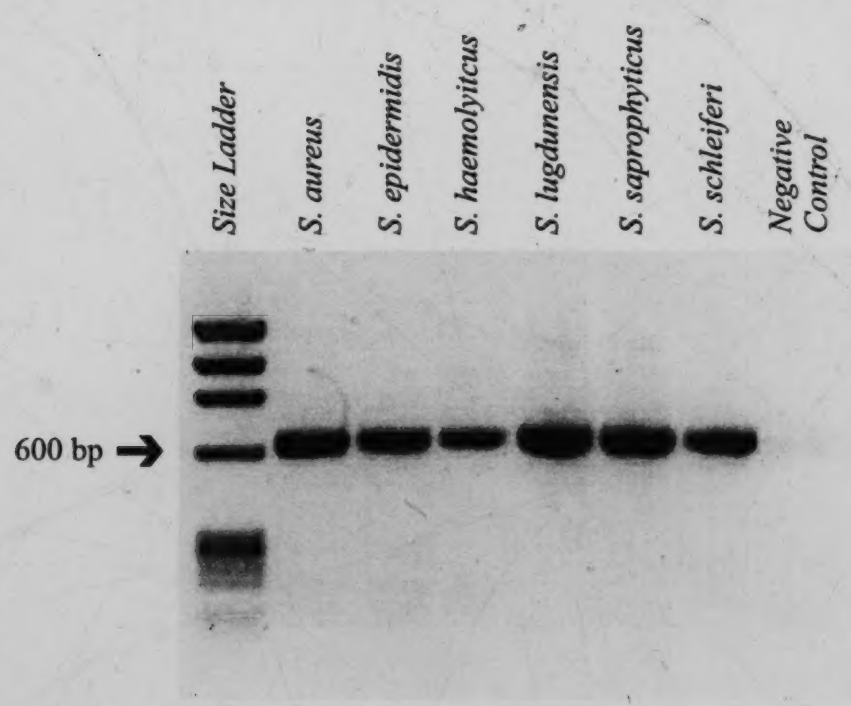


Figure 5.3. The HSP60 PCR of 6 *Staphylococcus* Species.
The 600 bp HSP60 PCR products from *S. aureus, S. epidermidis, S. haemolyticus, S. lugdunensis, S. saprophyticus,* and *S. schleiferi*. Run on 0.8% agarose gel, 1XTAE with ΦX174 RF DNA/*Hae*III digest size ladder.

100 µL of freshly thawed DH5α competent cells. The cells were carefully mixed and left on ice for 30 min. The bacterial cells were heat-shocked at 42°C for 90 s and then placed back on ice for 3 min. 800 µL of Luria broth (Becton Dickinson) was added to the cells, and the cells were grown on roller wheel for 45 min at 37°C. 100 µL of the culture was plated onto BHI agar plates with 100 µg/ml ampicillin (USB), 1000 µg IPTG (isopropylthio-β-D-galactoside, Gibo BRL) and 700 µg X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside, Gibco BRL). The plates were incubated at 37°C overnight.

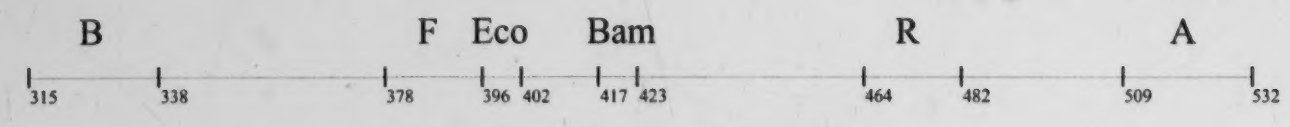
5.2.2.3. Boiling/Lysis AB PCR

A figure of the cloned HSP60 PCR product and multiple cloning site of pUC19 is shown in Figure 5.4.

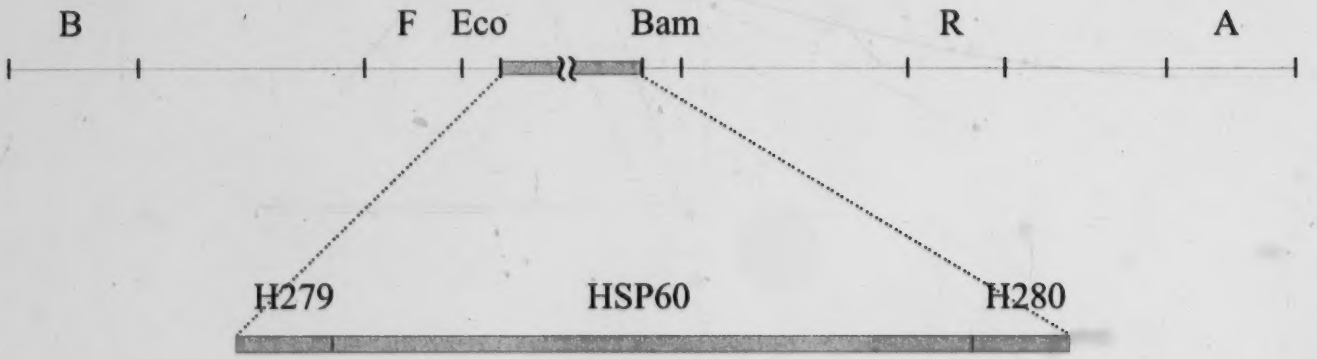
The colonies were screened using Pieter Roos' AB PCR, a modified version of Dynabeads Template Preparation Starter Kit protocol (Dynal, Great Neck, NY). The AB PCR primer binding sequences flanked the multiple cloning site. The primer sequences were 5'-biotin-GCTTCCGGCTCGTATGTTGTGTG-3' for the A1 primer and 5'-biotin-AAAGGGGGATGTGCTGCAAGGCG-3' for the B2 primer (University of Calgary Core DNA Lab, Calgary, Alberta). The A1 priming site was located at position 509 to 531 bp of pUC19. The B2 priming site was located at 315 to 337 bp. The PCR of the recombinant pUC19 clones resulted in an 807 bp product. pUC19 clones without an insert were 217 bp long.

A single white colony was selected from the agar plate. The colony was touched with a pipet tip and then dipped into 20 µl of water. The bacterial cells were lysed by boiling the water and the cellular debris was pelleted in 1 min at 13 000 rpm in the microcentrifuge. 10 µL of the supernatant was added to 38 ng of primer A1 and B2 (5 pmol), 10 nmoles dNTPs, 75 nmoles $MgCl_2$, 5 µl of 10×PCR reaction buffer and 2.5 U Taq. The final volume of the reaction was 50 µL. The thermal cycling conditions were: 30 s at 96°C, 25 cycles of 30 s at 94°C, 30 s at 55°C and 1 min at 72°C, and last cycle of 2 min at 72°C.

4 µL of AB PCR reaction was run with 2 µl of load dye on 0.8% agarose, 1×TAE gel at 100 V for 1 h. Positive clones (pUC19 with HSP60 PCR product inserted)



a) pUC19 multiple cloning site (from 315 to 532 bp)



b) pUC19 with HSP60 insert

Figure 5.4. The pUC19 Plasmid Multiple Cloning Site

a) The multiple cloning site of pUC19 before cloning. B and A are the AB PCR primer binding sites. F and R are the M13 forward and reverse sequencing primer binding sites respectively. Eco and Bam are the sites where the *Eco*RI and *Bam*HI restriction enzymes cut respectively. b) The multiple cloning site of pUC19 with the HSP60 PCR product inserted between the *Eco*RI and *Bam*HI cut sites. H279 and H280 are the HSP60 PCR primer sequences.

produced a 800 bp PCR product. The positive AB PCR products were cleaned with QIAquick PCR purification kit, resuspended in 10 mM Tris-HCl, pH=8.0, and the concentration measured with fluorimeter. The concentrations ranged from 50 to 100 ng/µl. The DNA from the AB PCR was used as a template for the cycle sequencing reaction described in Section 5.2.4.

5.2.3. Subcloning the Halved HSP60 PCR Products

A schematic of the sub-cloning HSP60 PCR products is shown in Figure 5.5. The AB PCR products of the positive HSP60 clones in pUC19 were 800 bp in size. The PCR products were digested with *Eco*RI and *Bam*HI to remove the sequencing primer sites, leaving a 600 bp product. The PCR fragments were digested roughly in half and blunt-ended by Klenow DNA polymerase. The fragments were ligated into *Sma*I digested pUC19 and transformed into *E. coli*. The antibiotic resistant and white colonies were AB PCR screened. Positive clones (pUC19 with halved HSP60 PCR product inserted) produced a ~500 bp PCR product.

5.2.3.1. AB PCR

The fragments were prepared by the previously described "boiling/lysis" AB PCR. 10 µL of the supernatant was mixed with 76 ng of primer A1 (10 pmol), 76 ng of primer B2 (10 pmol), 20 nmoles dNTPs, 150 nmoles $MgCl_2$, 10 µl of 10×PCR reaction buffer and 5 U Taq. The final volume of the reaction was 100 µL. 5 µL of AB PCR reaction was run on gel to verify PCR reaction.

5.2.3.2. *Eco*RI/*Bam*HI Digest

95 µL of AB PCR were double-digested with 10 U each of *Eco*RI and *Bam*HI in a final volume of 120 µL. The digests were incubated at 37°C for 1 h. Adding 300 µL of 98% ethanol and freezing at -20°C for 1 h precipitated the digests. The samples were spun at 13000 rpm in a microcentrifuge for 20 min. The ethanol was decanted and the DNA pellet was resuspended in 20 µL 10 mM Tris-HCl, pH=8.0. The entire digested PCR fragment solution was run on a 0.8% agarose, 1×TAE gel with 5 µL of uncut AB

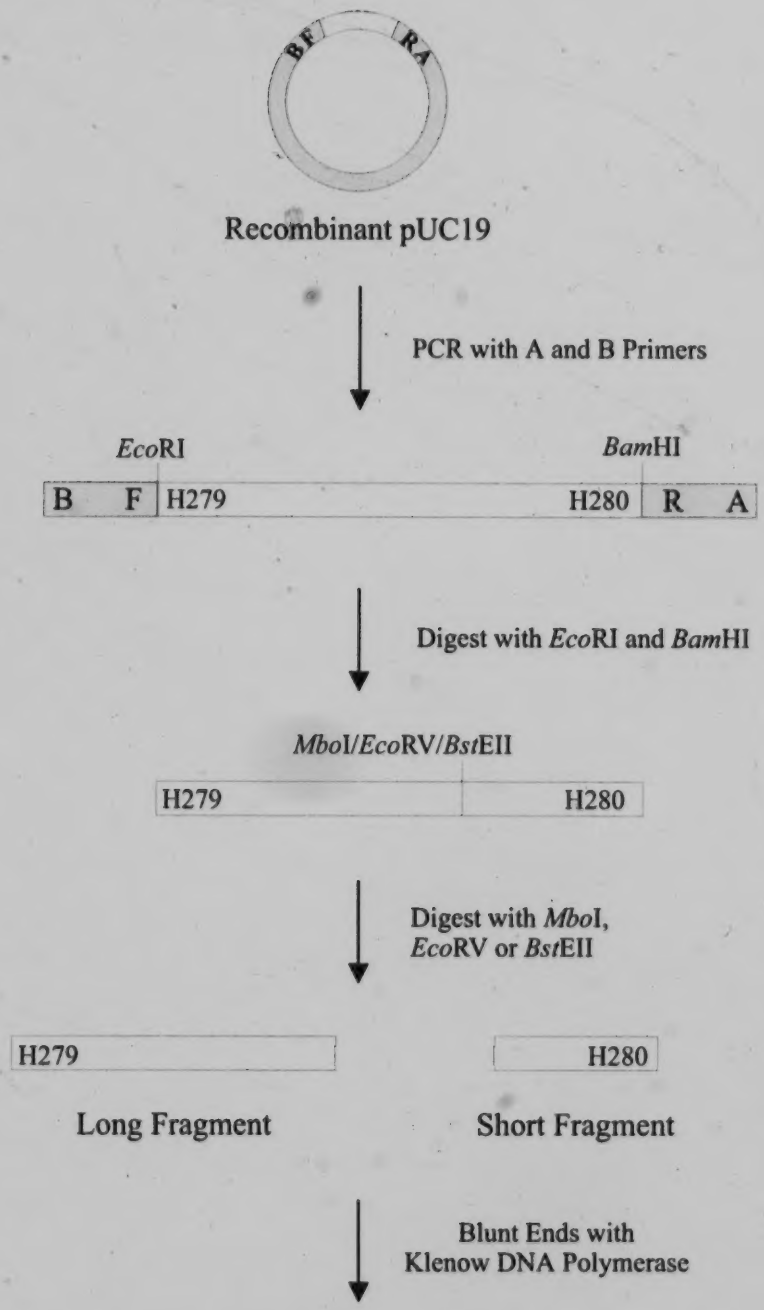


Figure 5.5. (Continued on next page)

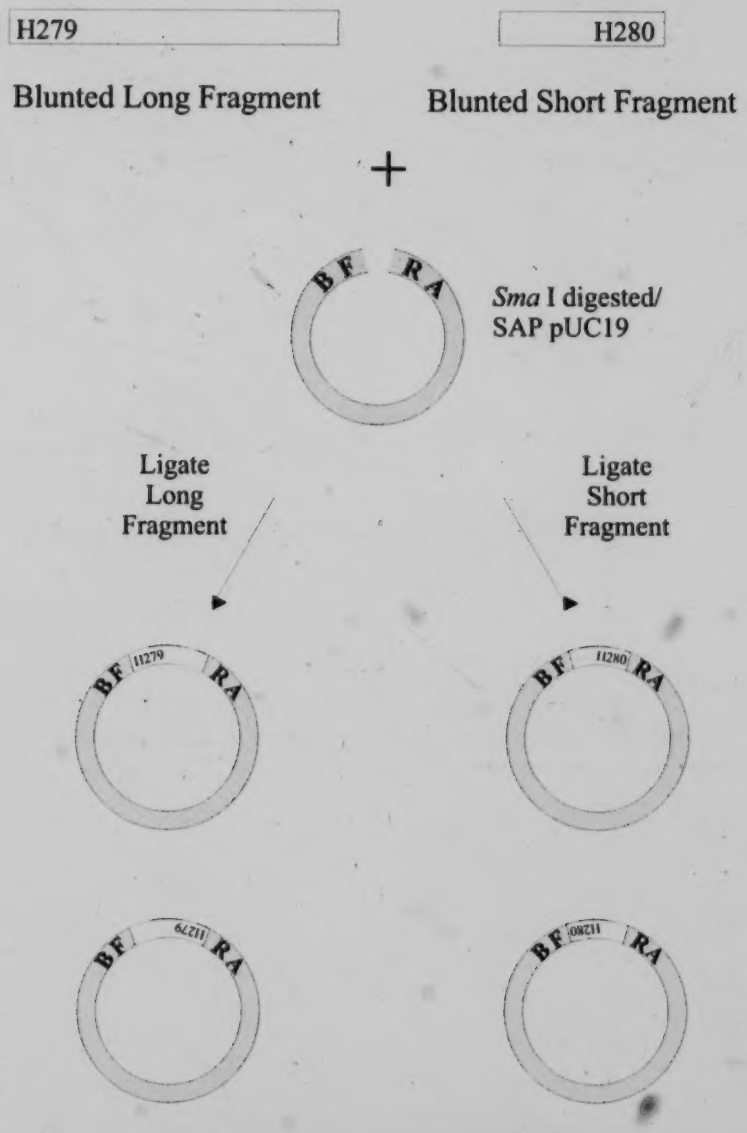


Figure 5.5. Subcloning the Halved HSP60 PCR Products.
The inserted HSP60 PCR product was amplified by PCR with A and B primers. The AB PCR product was cut with *Eco*RI and *Bam*HI restriction enzymes. The fragment was then cut approximately in half by one of three restriction enzymes: *Mbo*I, *Eco*RV or *Bst*EII. The cut pieces were separated on an agarose gel and the ends are blunted with Klenow DNA polymerase. The fragments were then cloned into *Sma*I digested pUC19 plasmid.

PCR fragment to verify complete cutting of *Eco*RI and *Bam*HI. The 600 bp bands were cut out of the gel and DNA was spun out of the gel slice using glass wool as a filter. The DNA was ethanol precipitated with 2.5 volumes (*i.e.* 2.5 × volume of DNA solution) of 98% ethanol, 1/10 volume of 3 M sodium acetate, pH=5.2, and freezing at -20°C for 1 h. The DNA was washed with 70% ethanol and air-dried.

5.2.3.3. Halving Digest

The HSP60 PCR products were halved into approximately 350 bp and 250 bp fragments, shown in Table 5.1.

S. aureus, S. epidermidis, S. saprophyticus and *S. schleiferi* AB PCR products were digested with 10 U of restriction enzyme *Mbo*I (Gibco BRL). The reactions were diluted to final volume of 20 µL and incubated at 37°C for 1 h. 2 U of Klenow DNA polymerase (Gibco BRL) was added and incubated at 37°C for 20 min. 1 µL of 10 mM dNTPs was added and the reaction was incubated at 37°C for 20 min. The enzyme was heat-killed at 75°C for 10 min. *S. haemolyticus* AB PCR product was digested with 10 U of restriction enzyme *Eco*RV (New England Biolabs, Inc., Beverly, MA). *S. lugdunensis* AB PCR product was digested with 10 U restriction enzyme *BstE*II (Gibco BRL). The reactions were diluted to final volume of 20 µL and incubated at 37°C for 1 h.

The digests were run on 1% agarose gel in 1×TAE. The long and short bands were cut out of the gel and DNA was spun out of the gel slice through glass wool. The fragments were extracted with 1 volume of buffer-saturated phenol (Gibco BRL). The aqueous layer was extracted with 1 volume of chloroform twice. The aqueous layer was ethanol precipitated and washed as previously described. The DNA was resuspended in TE buffer (10 mM Tris-Cl and 1 mM EDTA, pH=8.0) and the concentration measured with a fluorimeter. The concentrations ranged from 10-50 ng/µL.

5.2.3.4. Cloning the Halved HSP60 Gene Products into pUC19

1 µg pUC19 was digested with 10 U restriction enzyme *Sma*I. The digest was incubated at 37°C for 1 h. 3 U of shrimp alkaline phosphatase (SAP, Amersham Life Sciences, Buckinghamshire, England) was added and the reaction was incubated at 37°C

for 1 h. 1 µL of pUC19 was run on a gel to ensure complete cutting. The digested pUC19 was phenol/chloroform-extracted and aqueous layer ethanol precipitated as previously described.

The ligation reaction was prepared with 200 ng of the halved fragment, 80 ng of *Sma*I cut, SAP treated pUC19 (approximately 20:1 ratio of digest to pUC19) and 1 U of T4 DNA ligase (Gibco BRL). The total reaction volume was 15 µL. The ligation reactions were incubated at 16°C overnight. The recombinant pUC19 was heat-shock transformed into *E. coli* strain DH5α as previously described.

The white colonies were screened as previously described by AB PCR. Positive sizes of fragments are listed in Table 5.1. Positive clones were cycle sequenced as described in following Section 5.2.4.

Table 5.1. Size of Subcloned HSP60 PCR Fragments

	Size after halving digest		AB PCR Products in pUC19	
	Long (bp)	Short (bp)	Long (bp)	Short (bp)
S. aureus	342	275	559	492
S. epidermidis	342	275	559	492
S. haemolyticus	453	164	670	381
S. lugdunensis	349	268	566	485
S. saprophyticus	396	221	613	438
S. schleiferi	342	275	559	492

The table is a list of the expected insert sizes of the halved HSP60 gene sequences after Klenow end-polishing. The size of the long fragment includes the H279 primer. The size of the short fragment includes the H280 primer.

5.2.4. DNA Sequencing

The cycle sequencing reactions were prepared using Amersham RPN 2536 kit with ThermoSequenase DNA polymerase (Amersham Life Sciences) and dye-labeled primers (Kit 401487, PE Applied Biosystems, Foster City, CA). The recipe in Table 5.2 was used to prepare a single cycle-sequencing reaction. The four terminations were prepared in four separate tubes. The reagent mix contained the dNTPs, the enzyme,

buffer and the appropriate ddNTP. The templates used were AB PCR products from the original 600 bp HSP60 or halved HSP60 fragments cloned into pUC19.

Table 5.2. Dye-Labeled Primer Cycle-Sequencing Reaction

	A-termination	C-termination	G-termination	T-termination
Reaction ratio A:C:G:T	1	1	2	2
PCR template	30 ng	30 ng	60 ng	60 ng
M13 forward or reverse dye-primer	0.8 pmol (JOE)	0.8 pmol (FAM)	1.6 pmol (TAMRA)	1.6 pmol (ROX)
reagent mix	4 µL	4 µL	8 µL	8 µL
final volume	16 µL	16 µL	32 µL	32 µL

The thermal cycling conditions were: 30 cycles of 30 s at 95°C and 30 s at 55°C.

The 4 reaction tubes were combined and the sequencing samples were desalted and cleaned with the following protocol. 15 µL of streptavidin immobilized on 4% beaded agarose (Sigma, St. Louis, MO) was added and gently mixed for 20 minutes at room temperature in the dark. The streptavidin fixed the biotinylated AB PCR templates. The agarose beads were pelleted with a quick spin in the microcentrifuge and the supernatant was applied to a Microcon 30 (Amicon, Beverly, MA) membrane. The tubes were spun at 11000 rpm for 4-5 min with the microcentrifuge. The membrane was washed with 200 µL of 0.1×TE (1 mM Tris-HCl and 0.1 mM EDTA, pH=8.0). The tubes were spun again for 4-5 min at 11000 rpm. The tubes were inverted into a new tube and the cleaned DNA spun out of membrane after 4-5 min at 6000 rpm. The sample volumes ranged from 5 to 20 µL.

5.2.5. Gel Preparation

Capillaries were 50 µm ID, 150 µm OD, and typically 40 cm long (Polymicro Technologies, Phoenix, AZ). The inner walls of the capillaries were coated using Hjerten's method to covalently bind the polyacrylamide to the walls [27]. A solution of 20 µl of γ-methacryloxypropyltrimethoxysilane (Sigma) in 1 mL of 95% ethanol was drawn through the capillaries for 1 h.

To polymerize the gel in the capillaries, the injection ends were placed in a 50-ml 3-neck flask. The acrylamide solution was 6%T in 7 M urea, 0.6 g acrylamide and 4.2 g urea diluted to 10 ml with 1×TBE buffer (89 mM Tris, 89 mM boric acid and 2.0 mM ethylenediaminetetraacetic acid (EDTA), pH=8.0). The solution was passed through a 0.22 μm filter and added to the flask. The solution was degassed with a vacuum pump for 10 minutes. The polymerization of the acrylamide was initiated under an argon blanket by adding 10 μL of 10% w/v ammonium persulfate and 5 μL of TEMED (N,N,N',N'-tetramethylethylenediamine). The solution was immediately drawn through the capillaries by vacuum, with the detector ends of the capillaries under water. The vacuum was removed and the gel was left to polymerize overnight. Capillaries were visually inspected for bubbles before insertion into cuvette.

5.2.6. Sample Injection and Run

Sixteen gel-filled capillaries were inserted into the sheath-flow cuvette and pre-run at 150 V/cm with aligning solution of 0.4 nM JOE-labeled primer and 0.4 nM of ROX-labeled primer in 1×TBE buffer. Capillaries that generated no fluorescent spots or very large fluorescent spots were not run. After the alignment of the instrument was confirmed, the capillaries were heated to 55°C by setting the Peltier power supply at a constant current value of 2.5A. The aligning solution was then migrated from the capillary by electrophoresis of 1×TBE. At the desired temperature, the samples were injected onto the capillaries for 15-30 s at 100 V/cm. Samples were injected from a gold-plated microtitre plate, as previously described. The sample injection plate was replaced with a plate filled with gel and the electrophoresis was continued at 150 V/cm.

5.2.7. DNA Sequence Data Processing

The DNA sequences were determined by the base-calling algorithm BASS [28]. Alterations to the program for capillary sequencing were made by Stephen Lewis. The lane-finding routine was removed and the program was modified to accept the high sampling rate used by capillary electrophoresis. The program was also modified to compensate for mobility shifts.

The general binary data from a multiple capillary run was split into individual capillary electropherograms. An electropherogram was loaded into BASS. The capillary data was divided into eight windows, and the routine was applied to filter, mobility shift correct, and color separate the data. The program used a shortest path algorithm between the peak maxima to identify the bases. The data was edited in the sequence assembly program Sequencher V3.0 (Gene Codes Corp., Ann Arbour, MI). The consensus sequence was derived from sequencing both strands of the HSP60 insert with forward and reverse universal primers.

5.2.8. Phylogenetic Tree Construction

The public domain program ClustalW V1.7 was used to align the DNA sequences [29, 30]. The sequences were analyzed by programs in the PHYLIP (Phylogeny Inference Package) V3.5c Package [31, 32]. Distances between the six *Staphylococcus* species were estimated with the program DNADIST, using Kimura 2-correction parameter. A phylogenetic tree was constructed using the Neighbor-Joining method in the program NEIGHBOR. The stability of the tree was estimated by performing a bootstrap analysis with 100 replications using the programs SEQBOOT, DNADIST, NEIGHBOUR and CONSENSE. The phylogenetic trees were drawn using the DRAWTREE program. The HSP60 DNA sequences of closely related organisms were retrieved by a BLAST search of GenBank and EMBL data libraries [33].

5.2.9. Nucleotide Accession Numbers

Table 5.3. GenBank Accession Numbers

Partial HSP60 gene Sequences	Accession Number
S. aureus 8325-4	AF029244
S. epidermidis 9759	AF029245
S. haemolyticus ATCC 29970	U92809
S. lugdunensis CRSN 850412	U92810
S. saprophyticus KL122	AF029246
S. schleiferi ATCC 43808	AF033622

Table 5.3 lists the GenBank Accession numbers for the sequences from this thesis. The sequences were presented 5' to 3' from first base after the H279 primer to the base before the sequence complementary to the H280 primer.

5.3. RESULTS AND DISCUSSION

5.3.1. HSP60 Gene Sequences

5.3.1.1. Phylogenetic Tree

S. aureus, *S. epidermidis*, *S. haemolyticus*, *S. lugdunensis*, *S. saprophyticus*, and *S. schleiferi* partial HSP60 gene sequences that were determined in this thesis are found in Appendix II. The BLAST search of GenBank for sequences similar to the HSP60 (a.k.a. GroEL) gene generated a large list of organisms with high sequence similarity. This result was not surprising considering the ubiquitous nature of the heat shock protein. The ten most related organisms from the BLAST search are listed in Table 5.4. The first three are *Staphylococcus* species and the other seven are Low G+C, gram-positive bacteria of various genera.

Table 5.4. Related Organisms from BLAST

Organism	Accession Number
Staphylococcus hyicus	AF019777
Staphylococcus delphini	AF019774
Staphylococcus intermedius	AF019773
Bacillus stearothermophilus	L10132
Bacillus subtilis	M81132
Thermoanaerobacter brockii	U56021
Clostridium perfringens	X62914
Clostridium acetobutylicum	M74572
Lactobacillus lactis	X71132
Lactobacillus zeae	AF010281

My data for *S. aureus*, *S. epidermidis*, *S. haemolyticus*, *S. lugdunensis*, *S. saprophyticus*, and *S. schleiferi* HSP60 gene sequences showed nucleotide similarity from 81 to 87%. The sequence similarity values between pairs of species are listed in Table 5.5 in the top right portion of the table. The bottom left portion represents the genetic distance between two species, where zero distance separates two identical objects and the

TABLE 5.5. Staphylococcus HSP60 Gene Genetic Distances and Sequence Similarities (%)

	S.aureus	*S.haemolyticus*	*S.epidermidis*	*S.lugdunensis*	*S.saprophyticus*	*S.schleiferi*
S.aureus		87%	84%	85%	84%	82%
S.haemolyticus	0.14		83%	85%	83%	83%
S.epidermidis	0.19	0.20		83%	83%	81%
S.lugdunensis	0.17	0.17	0.19		84%	81%
S.saprophyticus	0.18	0.20	0.20	0.18		82%
S.schleiferi	0.20	0.19	0.23	0.22	0.21	

Sequence Similarities (%): Top Right Corner
Genetic Distance: Bottom Left Corner
(Data generated from this thesis)

TABLE 5.6. HSP60 Gene Genetic Distances and Sequence Similarities (%)

	C.perfringens	*C.acetobutylicum*	*T.brockii*	*B.stearothermophilus*	*B.subtillis*	*S.schleiferi*	*S.hyicus*	*S.delphini*	*S.intermedius*	*S.aureus*	*S.haemolyticus*	*S.lugdunensis*	*S.epidermidis*	*S.saprophyticus*	*L.lactis*	*L.zeae*
C.perfringens		78%	72%	68%	65%	67%	66%	63%	63%	68%	68%	67%	67%	66%	66%	63%
C.acetobutylicum	0.28		73%	69%	69%	65%	65%	62%	63%	65%	65%	66%	66%	65%	65%	66%
T.brockii	0.38	0.35		70%	68%	66%	66%	64%	65%	66%	65%	67%	66%	68%	67%	64%
B.stearothermophilus	0.44	0.44	0.41		77%	70%	69%	70%	71%	69%	71%	69%	68%	69%	65%	68%
B.subtillis	0.50	0.43	0.45	0.28		68%	66%	65%	66%	66%	64%	66%	66%	66%	65%	66%
S.schleiferi	0.50	0.54	0.52	0.41	0.46		86%	84%	84%	82%	83%	81%	81%	82%	68%	65%
S.hyicus	0.51	0.53	0.51	0.43	0.48	0.15		83%	84%	81%	80%	81%	80%	80%	64%	66%
S.delphini	0.60	0.64	0.57	0.41	0.51	0.18	0.19		94%	82%	82%	79%	78%	80%	61%	63%
S.intermedius	0.60	0.60	0.54	0.40	0.49	0.18	0.18	0.06		82%	82%	80%	79%	81%	63%	65%
S.aureus	0.49	0.55	0.52	0.43	0.50	0.20	0.23	0.21	0.20		87%	85%	84%	84%	64%	65%
S.haemolyticus	0.49	0.55	0.53	0.40	0.53	0.19	0.24	0.21	0.20	0.14		85%	83%	83%	60%	65%
S.lugdunensis	0.51	0.52	0.50	0.43	0.50	0.22	0.22	0.24	0.23	0.17	0.17		83%	84%	65%	64%
S.epidermidis	0.51	0.51	0.53	0.46	0.49	0.23	0.24	0.26	0.25	0.19	0.20	0.19		83%	64%	64%
S.saprophyticus	0.52	0.54	0.48	0.44	0.50	0.21	0.24	0.24	0.23	0.18	0.20	0.18	0.20		67%	65%
L.lactis	0.51	0.53	0.50	0.51	0.53	0.47	0.51	0.57	0.52	0.50	0.56	0.55	0.54	0.49		68%

Sequence Similarities (%): Top Right Corner, Genetic Distances: Bottom Left Corner

most divergent species are separated by infinite distance [4]. Of the six studied species, the most closely related are *S. aureus* and *S. haemolyticus* with 87% sequence similarity and a distance of 0.14. The least related species appears to be *S. epidermidis* and *S. schleiferi* with the lowest sequence similarity of 81% and largest distance of 0.23.

An estimated phylogenetic tree of the *Staphylococcus* species studied in this thesis was constructed with the three *Staphylococcus* species retrieved from the BLAST search (Figure 5.6). Another phylogenetic tree was constructed with the other organisms from the BLAST search (Figure 5.7). It was not surprising to observe that organisms within a genus are most related to each other. The trees were unrooted because the common ancestor of the species is not known. The lengths of the branches are roughly drawn to scale, illustrating the genetic distances between the species. The sequence similarities and distance measurements for all 16 sequences are shown in Table 5.6. It appears that the most related species are *S. delphini* and *S. intermedius*.

The trees constructed here infer gene phylogeny and not species phylogeny, because this tree compared the HSP60 gene sequences [34]. Polymorphisms exist within a species, and the polymorphisms may have occurred a long time ago. As a result the phylogenetic history derived from genes may be different from the species that produce these genes. The phylogenetic trees presented here are qualitative and estimate the relationships between the species. The phylogeny between the trees may not be important for the *Staphylococcus* species identification in a diagnostic lab. Bacterial identification will probably rely mostly on the sequence similarity. However, the tree phylogeny may be important in identifying unknown species and in studying the evolution of bacterial species.

There was only one other HSP60 gene sequence in GenBank that could be used to compare within-strain sequence variation. The sequence was *S. aureus* strain 912 (Accession # D14711), which is a methicillin antibiotic-resistant *Staphylococcus*. The Clustal DNA alignment of the two species is shown in Figure 5.8. There is 96% sequence similarity between *S. aureus* strain 912 and *S. aureus* strain 8325-4. The biggest difference between the two sequences is a 9-base gap at position 257-265 bases

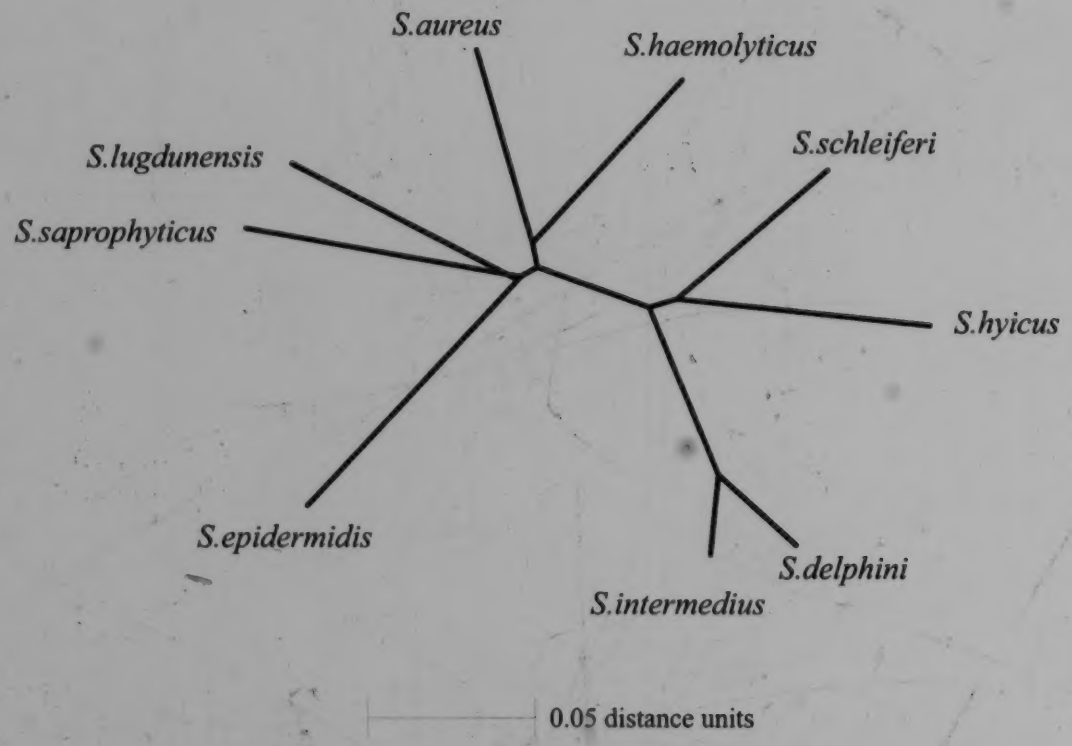


Figure 5.6. Unrooted Phylogenetic Tree of *Staphylococcus* HSP60 Gene Sequences.

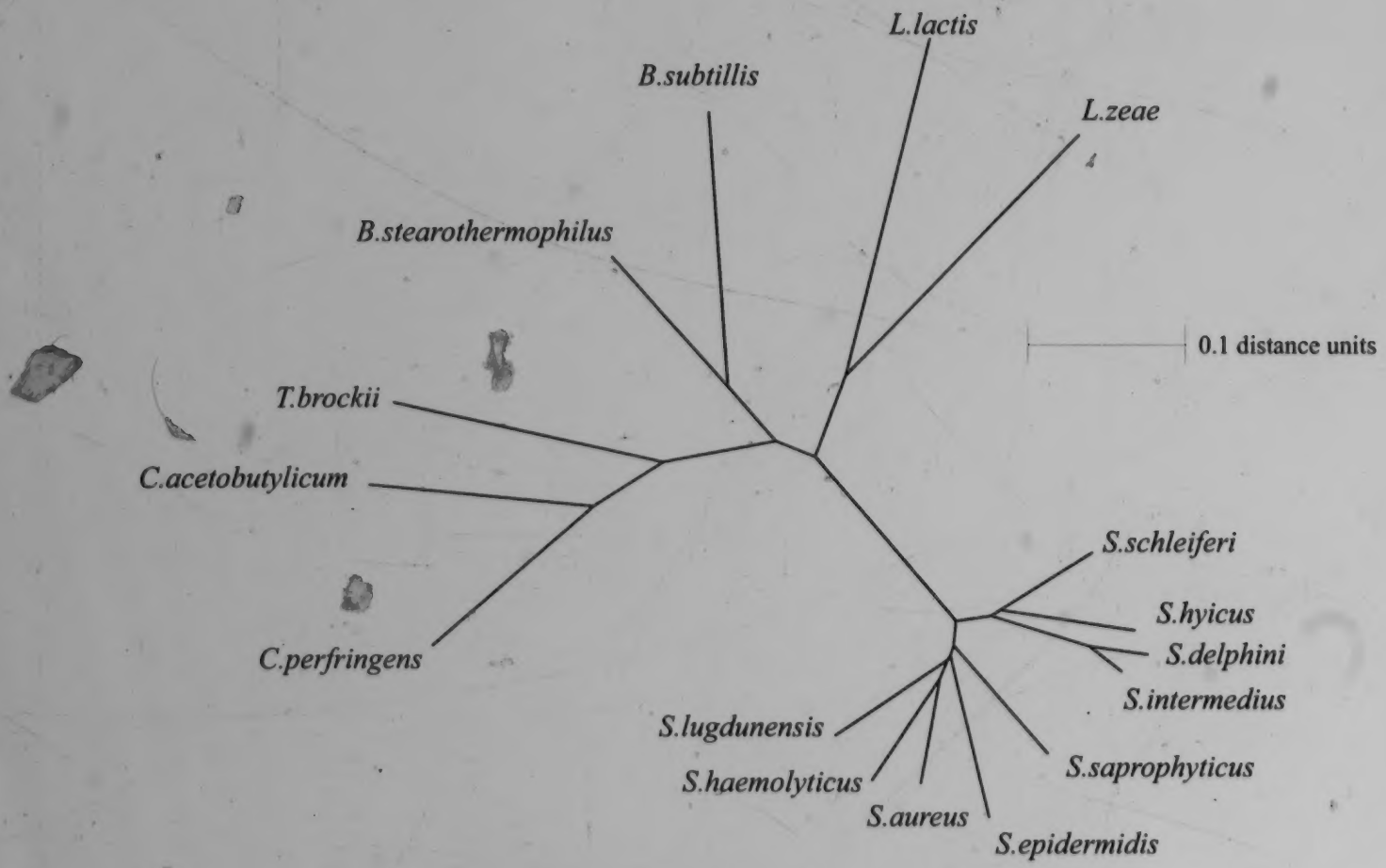


Figure 5.7. Unrooted HSP60 Gene Phylogenetic Tree of *Staphylococcus* and Other Closely Related Organisms From BLAST Search.

Figure 5.8. DNA Sequence Alignment of *S. aureus* 8325-4 and *S. aureus* 912
Nucleotide sequence agreements are indicated by "*" underneath. Sequence disagreements are unmarked. Gaps in sequences are indicated by "-".

```
8325-4     1 ATVLAQAMIQEGLKNVTSGANPVGLRQGIDKAVKVAVEALHENSQKVENKNEIAQVGAIS
912          ATVLAQAMIQEGLKNVTSGANPVGLRQGIDKAVKVAVEALHENSQKVENKNEIAQVGAIS
             ************************************************************

8325-4    61 AADEEIGRYISEAMEKVGNDGVITI---EESNGLNTELEVVEGMQFDRGYQSPYMVTDSD
912          AADEEIGRYISEATEKVGNDGVITIITIEESNRLNTELEL--GMQFDRGYQSPYMVTDSD
             ************* ***********    **** ******:  ******************

8325-4   121 KMVAELERPYILVTDKKISSFQDILPLLEQVVQSNRPILIVADEVEGDALTNIVLNRMRG
912          KMVAELERPYILVTDKKISSFQDILPLLEQVVQSNRPILIVADEVEGDALTNIVLNRMRG
             ************************************************************

8325-4   181 TFTAVAV
912          TFTAVAV
             *******
```

Figure 5.9. Amino Acid Sequence Alignment of *S. aureus* 8325-4 and *S. aureus* 912
Amino acid sequence agreements are indicated by "*" underneath. Amino acid disagreements are unmarked. Gaps in sequences are indicated by "-".

and a 6-base insertion at position 303-308 in strain 8325-4. The gap represents a loss of 3 amino acids and the insertion represents an extra 2 amino acid in strain 8325-4 compared to strain 912 (Figure 5.9). These differences may be used to identify different strains of *Staphylococcus* within a species, which could be important in clinical laboratories.

5.3.1.2. Cloning HSP60 PCR Products

The DNA sequences of the six *Staphylococcus* species start a library of reference sequences to identify bacteria by the HSP60 gene sequencing. Since the method of subcloning was very laborious, it is impractical for a larger number of reference sequences. It was assumed that in any force-cloning method, the recognition sequence for the restriction enzymes used was not in the region of interest. Blunt-end cloning of HSP60 PCR products would avoid this problem at the cost of lower efficiency of ligation.

The ideal strategy for HSP60 microbial identification diagnostics is to perform PCR analysis with the HSP60 primers from a specimen (e.g. a blood sample) collected from the patient. The PCR product would be sequenced directly using the HSP60 PCR primers and dye-labeled dideoxynucleotides. The sequences would then be compared to the reference library and the bacterium identified. The average read-lengths of dye-terminators are approximately 300-350 bases for ds plasmids and ss templates [35]. I am not certain if the entire 600 bp would need to be sequenced. More sequences must be placed in a database to determine the minimum amount of DNA that needs to be sequenced to correctly identify the bacteria. Most likely, sequencing from both ends of the PCR template will give enough data to identify the bacterial species.

Another slight variation of the above approach is to amplify the HSP60 PCR product with modified H279 and H280 primers. These primers would contain the binding sites for forward or reverse universal primers, illustrated in Figure 5.10. After PCR with these primers, the PCR product could be sequenced with the forward and reverse primers, assuming the presence of only one type of PCR product. The advantages of this modification are the use of dye-labeled primers for longer sequencing read length of 500 bases [35]. Also, the primers do not have degenerate sites, allowing for higher

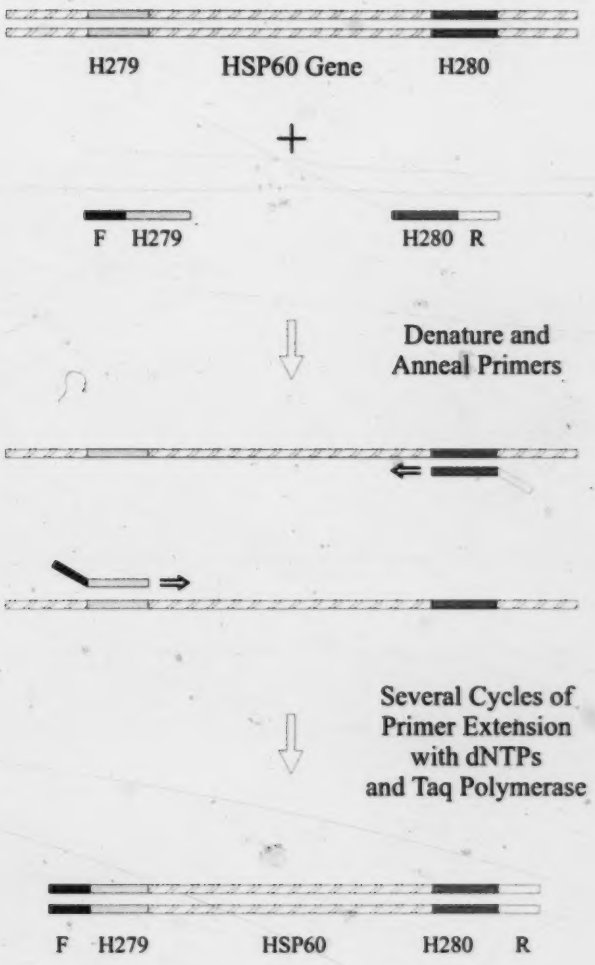


Figure 5.10. HSP60 PCR with Modified PCR Primers

The HSP60 gene is amplified with modified HSP60 PCR primers, which have forward (F) and reverse (R) universal priming sites. The final PCR product can be directly sequenced with dye-labelled sequencing primers.

annealing temperatures for the cycle sequencing reaction (37°C *vs.* 55°C). The higher temperature would reduce non-specific binding and reduce ghost peaks in the sequencing electropherogram.

It remains to be seen how the HSP60 diagnostic would be implemented in a clinical lab. The ability to perform PCR directly from a patient's specimen needs to be investigated. There are publications that report detecting and identifying bacterial infections by PCR from blood [36-38], cerebrospinal fluid [39], sputum [40, 41], and urine specimens [42]. The possible inhibition of PCR reaction needs to be studied as well as the specificity of the HSP60 PCR primers for bacterial DNA. The possibility of a multiple infection with several bacteria strains and/or species will require further study. Possibly, an initial culture step will be required to isolate the different bacteria before the HSP60 PCR can be performed.

5.3.2. Capillary DNA Sequencing

Preparation of ss templates is laborious but results in good quality sequence data. PCR templates are very quick and easy to prepare. A protocol has been developed to sequence double-stranded PCR products successfully as shown in Figure 5.11.

The previous exercise to sequence the wt M13 clones suggested that the PCR templates required rigorous cleaning. Adding a Qiagen PCR purification step before the cycle-sequencing reduced the huge fluorescent signals that used to plague the wt M13 sequencing runs (Figure 7.1). However the Qiagen clean-up did not remove the stretches that were visible in sequencing runs at 35-40°C. With a new Peltier heating device, and more careful control of the temperature at 55°C, the stretches and lumps were no longer observed when using PCR templates, as demonstrated in Chapter 3.

It is unknown what caused the stretch sample artifacts. The stretches occur at different base positions and differing sequences. Most likely these artifacts were caused by secondary structure that melted at 55°C. It may be intra-strand interaction or interaction with the sequencing template. The latter seems more probable because these

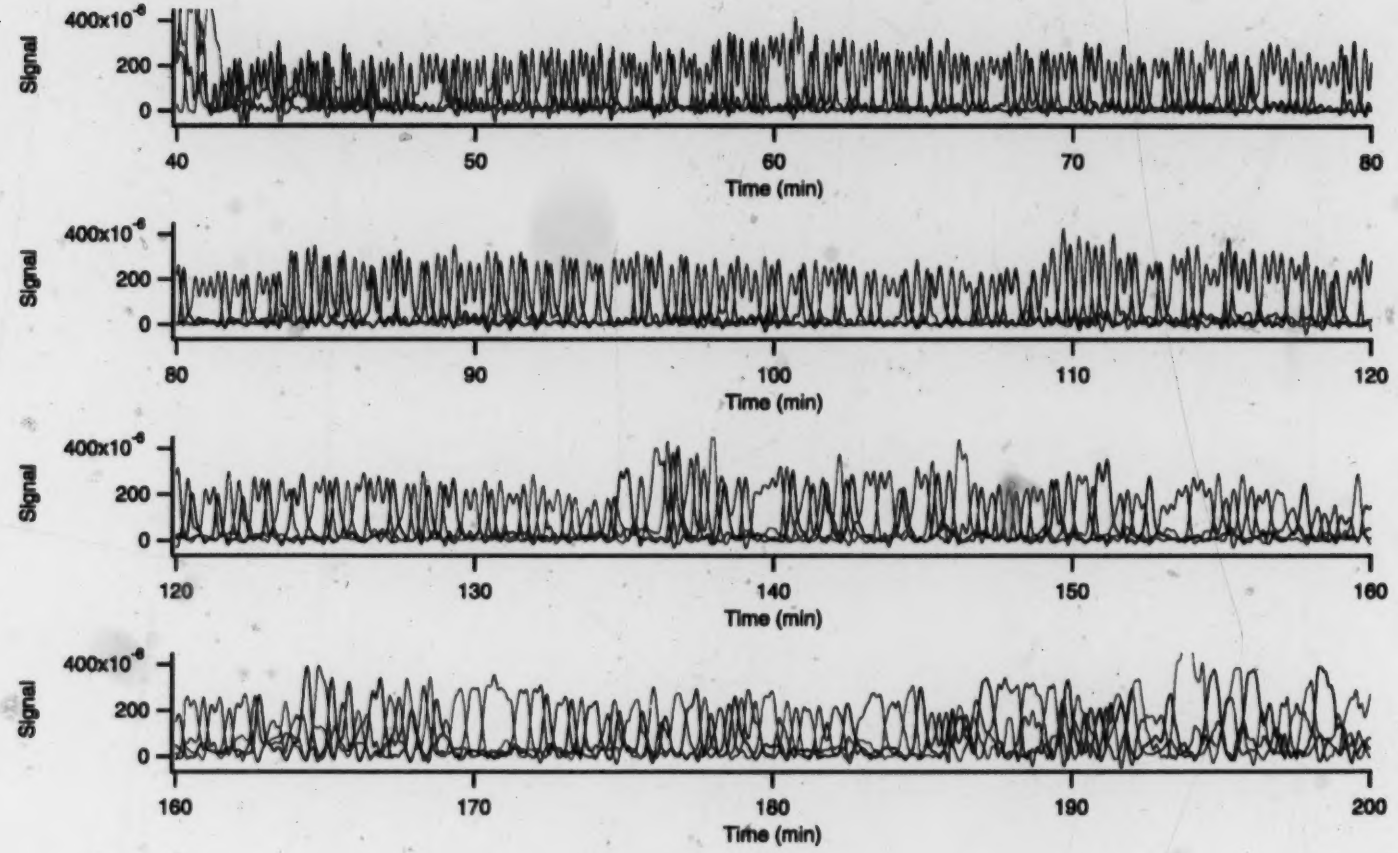


Figure 5.11. The HSP60 Gene Sequence of *S. epidermidis.*
Forward-primed reaction of *S.epidermidis*. Run on 6%T polyacrylamide gel at 55°C and 150 V/cm.

artifacts were also observed for ds plasmid templates and not observed when sequencing ss templates.

Fourteen runs were required on the 16-capillary instrument to sequence the six *Staphylococcus* species. Of the 224 capillaries, 10% of the capillaries had bubbles from the polymerization step and 40% of the capillaries were unusable because there were no sequencing peaks, or the peaks were poorly resolved. In some cases the peaks were so poorly resolved that the fragments co-migrated as a large peak. The unusable capillaries were associated with the capillary current dropping after sample injection or the result of poor gel and gel coatings. About 50% of the capillaries, about 110 capillaries, were successful in the sense that sequencing peaks were observed. Of the successful capillaries, 41 capillary runs (37%) had sufficient signal levels and peak resolution to be submitted to the BASS base-calling program. Of all the runs, 35 of the successful capillaries came from only 5 runs. The results indicate an overall capillary success rate of 18%. While this success rate is an improvement from the 6% success rate from the wt M13 runs (Section 2.3.3.2), the variability of the gels is unacceptable.

There are many intertwined parameters that affect DNA sequencing with polyacrylamide gels in capillaries. These parameters include the examination of the amount of DNA loaded onto the capillary, the effect of heating of the capillaries and the variability of gel polymerization and the gel-coating chemistries to the capillary walls. A study to quantify the optimum amount of DNA that can be loaded onto a gel-filled capillary is necessary, evident by the capillary current drop after a sample injection. This current drop can be very sudden suggesting that either a bubble is formed at the capillary tip or too much DNA has been loaded onto the capillary [43].

The temperature precision is important in gel resolution for DNA sequencing [44]. In this study without the PID control on the capillary heaters, there was usually a temperature drift of about 0.5°C over the course of the 3 h run. While the drift rate is very slow, it is likely to significantly affect the sequencing resolution.

The variability of the gel chemistries can be addressed by using a replaceable gel matrix, which will reduce differences in the gels from batch to batch. Studies using replaceable gels on single capillary instruments have shown the loss of sequencing

resolution due to the failure of capillary coating [44]. The search for a gel coating that is stable with heating and repeated gel replacement is on-going.

A remarkable solution to the gel problem is PE Applied Biosystems' poly-N,N-dimethylacrylamide (DMA) gel [45]. The DMA gels separate 600 nucleotides with 98.5% accuracy in under 3 h, heated to 50°C and used to refill capillaries at least 100 times [46]. The DMA polymer does not require a coated capillary as it "dynamically" coats the glass walls to suppress electroosmotic flow and it acts as the sieving matrix. DMA gels will solve the gel problem with its low viscosity for fast refilling of capillaries and its dynamic wall coating. Unfortunately, this material was not commercialized in time to be used in my experiments.

5.4. CONCLUSIONS

This chapter illustrated the ability of the 16-capillary DNA sequencer to analyze PCR template samples. The sequence similarity of the HSP60 gene sequences ranged from 81 to 87%, which appeared to be sufficient to identify different *Staphylococcus* species. The HSP60 gene sequences were submitted to GenBank to create a library to identify bacteria based on this highly conserved gene. A proposed phylogenetic tree was constructed for the *Staphylococcus* species sequenced in the this thesis and some sequences retrieved from a BLAST search.

Although high sample throughput was demonstrated, a small fraction of the electropherograms were of value. As a result a large number of runs were required to sequence a small number of templates. The biggest problem apparently does not lie in the instrumentation but instead with the separation chemistry, which has been a limitation to capillary DNA sequencing. An improved polymer, with low viscosity, will increase the success rate of capillaries, and ultimately make multiple capillary instruments truly high throughput.

5.5. REFERENCES

(1) Weir, B. S. *Genetic Data Analysis II*, second ed.; Sinauer Associates, Inc.: Sunderland, MA, **1996**.

(2) Hillis, D. M.; Moritz, C., Eds. In *Molecular Systematics*; Sinauer Associates, Inc.: Sunderland, MA, **1990**, pp 588.

(3) Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, **1987.**

(4) Swofford, D. L.; Olsen, G. J. In *Molecular Systematics*; Hillis, D. M., Moritz, C., Eds.; Sinauer Associates, Inc.: Sunderland, MA, **1990**, pp 411-501.

(5) Davison, D. *Bull. Math. Biol.* **1985,** *47*, 437-474.

(6) Needleman, S. B.; Wunsch, C. D. *Journal of Molecular Biology* **1970,** *48*, 443-453.

(7) Kruskal, J. B. In *Time Warps, String Edits, and Macromolecules*; Sankoff, D., Kruskal, J. B., Eds.; Addison-Welsey: Reading, MA, **1983**, pp 1-44.

(8) Smith, T. F.; Waterman, M. S. *Journal of Molecular Biology* **1981,** *147*, 195-197.

(9) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. *Journal of Molecular Biology* **1990,** *215*, 403-410.

(10) Dayhoff, M. O.; Schwartz, R. M.; Orcutt, B. C. In *Atlas of Protein Sequence and Structure*; National Biomedical Research Foundation: Washington, D.C., **1979**; Vol. 5, Supplement 3, 1978, pp 345-351.

(11) Henikoff, S.; Henikoff, J. G. *Proc. Natl. Acad. Sci. USA* **1992,** *89*, 10915-19.

(12) Karlin, S.; Altschul, S. F. *Proc. Natl. Acad. Sci. USA* **1990,** *87*, 2264-2268.

(13) Karlin, S.; Altschul, S. F. *Proc. Natl. Acad. Sci. USA* **1993,** *90*, 5873-7.

(14) Jukes, T. H.; Cantor, C. R. In *Mammalian Protein Metabolism*; Munro, H. N., Ed.; Academic Press: New York, **1969**, pp 21-123.

(15) Kimura, M. *Proc. Natl. Acad. Sci. U.S.A.* **1980,** *78*, 454-458.

(16) Saitou, N.; Nei, M. *Mol. Biol. Evol.* **1987,** *4*, 406-425.

(17) Felsenstein, J. *Journal of Molecular Evolution* **1981,** *17*, 368-376.

(18) Felsenstein, J. *Evolution* **1985,** *39*, 783-791.

(19) Pascual, C.; Lawson, P. A.; Farrow, J. A. E.; Navarro Giminez, M.; Collins, M. D. *International Journal of Systematic Bacteriology* **1995,** *45*, 724-728.

(20) Lawson, P. A.; Falsen, E.; Akervall, E.; Vandamme, P.; Collins, M. *International Journal of Systematic Bacteriology* **1997,** *47*, 899-903.

(21) Foster, G.; Ross, H. M.; Hutson, R. A.; Collins, M. D. *International Journal of Systematic Bacteriology* **1997,** *47*, 724-726.

(22) Balkwill, D. L.; Reeves, R. H.; Drake, G. R.; Reeves, J. Y.; Crocker, F. H.; Baldwin King, M.; Boone, D. R. *FEMS Microbiological Reviews* **1997,** *20*, 201-216.

(23) Smida, J.; Kazda, J.; Strackebrandt, E. *International Journal of Leprosy* **1988,** *56*, 449-454.

(24) Vandamme, P.; Pot, B.; Gillis, M.; De Vos, P.; Kersters, K.; Swings, J. *Microbiological Reviews* **1996,** *60*, 407-438.

(25) Canhos, V. P.; Manfio, G. P.; Blaine, L. D. *Antonie van Leeuwenhoek* **1993,** *64*, 205-229.

(26) Goh, S. H.; Potter, S.; Wood, J. O.; Hemmingsen, S. M.; Reynolds, R. P.; Chow, A. W. *Journal of Clinical Microbiology* **1996,** *34*, 818-823.

(27) Hjerten, S. *Journal of Chromatography* **1985,** *347*, 181-198.

(28) http://www-genome.mi.mit.edu/ftp/pub/software/Bass/, Whitehead Institute.

(29) Thompson, J. D., Higgins D. G., Gibson, T. J. *Nucleic Acids Research* **1994,** *22*, 4673-4680.

(30) Higgins, D. G., Thompson, J.D., Gibson, T. J. *Methods in Enzymology* **1996,** *266*, 383-402.

(31) Felsenstein, J. *Cladistics* **1989,** *5*, 164-166.

(32) Felsenstein, J., PHYLIP (Phylogeny Inference Package) version 3.5c., Department of Genetics, University of Washington, **1993**.

(33) http://www.ncbi.nlm.nih/gov, GenBank, National Center for Biotechnology Information.

(34) Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, **1987**.

(35) Naeve, C. W.; Buck, G. A.; Niece, R. L.; Pon, R. T.; Robertson, M.; Smith, A. J. *BioTechniques* **1995,** *19*, 448-452.

(36) Rattanathongkom, A.; Sermswan, R. W.; Wongratanacheewin, S. *Molecular & Cellular Probes* **1997,** *11*, 25-31.

(37) Newcombe, J.; Cartwright, K.; Palmer, W. H.; McFadden, J. *Journal of Clinical Microbiology* **1996,** *34*, 1637-40.

(38) Zhang, Y.; Isaacman, D.; Wadowsky, R. M.; Rydquist-White, J.; Post, J. C.; Ehrlich, G. D. *Journal of Clinical Microbiology* **1995,** *33*, 596-601.

(39) Prariyachatigul, C.; Chaiprasert, A.; Meevootisom, V.; Pattanakitsakul, S. *Journal of Medical and Veterinary Mycology* **1996,** *34*, 251-8.

(40) Karpati, F.; Jonasson, J. *Molecular & Cellular Probes* **1996,** *10*, 397-403.

(41) Campbell, P. W.; Phillips, J. A.; Heidecker, G. J.; Krishnamani, M. R.; Zahorchak, R.; Stull, T. L. *Pediatric Pulmonology* **1995,** *20*, 44-9.

(42) van Vollenhoven, P.; Heyns, C. F.; de Beer, P. M.; Whitaker, P.; van Helden, P.; Victor, T. *Urological Research* **1996,** *24*, 107-11.

(43) Swerdlow, H.; Dew-Jager, K. E.; Brady, K.; Grey, R.; Dovichi, N. J.; Gesteland, R. *Electrophoresis* **1992,** *13*, 475-83.

(44) Voss, K. O. Ph.D., University of Alberta, Edmonton, Alberta, **1998**.

(45) Rosenblum, B. B.; Oaks, F.; Menchen, S.; Johnson, B. *Nucleic Acids Research* **1997,** *25*, 3925-3929.

(46) ABI PRISM 310 Genetic Analyzer: Performance Optimized Polymers, Perkin-Elmer/Applied Biosystems, **1997**.

Chapter 6: Accuracy of Capillary Sequencing

6.1. INTRODUCTION

With high throughput sequencing projects, manual calling of DNA sequences is impractical. We used a modified version of BASS from MIT-Whitehead group to call the DNA sequences. It was our group's first implementation of a computer base-calling program to call large amounts of DNA sequences. It was our experience that extensive manual editing of data was required as the program was not perfect for our use. This chapter presents a study of BASS software for base calling and accuracy of capillary sequencing and a comparison with PE Applied Biosystems (ABI) slab gel sequencers.

6.1.1. Background

A brief description of the BASS base-calling software can be found in Section 5.2.7.

The spectral overlap is generated when the fluorescence from a single dye is collected from more than one bandpass filter. An example of this overlap is ROX's emission spectrum passing through the ROX bandpass filter (610DF10) and the TAMRA bandpass filter (580DF10). The color-separation algorithm, in BASS's base-calling program, used a matrix algorithm to reduce the spectral overlap for each of the dyes studied here. In the data output, FAM (C-termination) was a blue trace, JOE (A-termination) was a green trace, TAMRA (G-termination) was a yellow or black trace and ROX (T-termination) was a red trace. A poorly separated color matrix created spurious peaks, which resulted in an incorrect base assignment.

A mobility shift resulted from differences in electrophoretic mobility of the fluorescent dyes attached to the sequencing primer. Mobility shifts were more pronounced for shorter DNA fragments, where the dye comprised a greater percentage of the overall mass. Mobility shifts were more severe for dye-labeled primers than dye-terminator labeled DNA fragments. For the dye-primers used in this study, TAMRA (G-terminated) and ROX (T-terminated) created the largest mobility shifts. G-terminated fragments migrated quickly compared to A- and C-terminated fragments. Following an A- or C-terminated fragments, G-terminated fragments often co-migrated and overlapped

the proceeding peak (Figure 6.1). T-terminated fragments tended to run slower than A- and C-terminated fragments. T-terminated fragments migrated about half a base slower before an A-terminated fragments and a full base slower before a C-terminated fragments (Figure 6.2).

The accuracy of the capillary DNA sequencing was evaluated by comparing individual runs to the consensus sequence. The consensus sequence was compiled from sequencing both strands of the HSP60 gene sequence. The errors were identified as deletion (Del), insertion (Ins), miscall (Mis), and no-call (NoC). A deletion error occurred when the base-calling program or editor missed a base in the run that was present in the consensus sequence. An insertion error occurred when program or editor added a base to the sequence that was not present in the consensus sequence. A miscall occurred when the program or editor identified a base incorrectly compared to the consensus sequence. A no-call occurred when the program recognized a base was present but could not determine the identity of that base.

6.2. EXPERIMENTAL

6.2.1. DNA Sequencing Data

The errors from capillary sequencing were analyzed using the sequencing runs from the original cloned 600 bp HSP60 PCR products and not the sub-cloned halved HSP60 PCR products. The cloned HSP60 samples and dye-labeled primer sequencing reactions were described in Section 5.2.4. A total of 26 runs were used in this analysis: 14 forward-primed reactions and 12 reverse-primed reactions. Twenty-one runs were run on the 16-capillary sequencer as described in Section 5.2.6. Five runs were from the 5-capillary instrument, using similar conditions outlined in Section 2.2.3.2.

6.2.2. Data Processing and Editing

An individual electropherogram was loaded into the BASS program. Once BASS was finished processing the data, the ".sep" files, files which were smoothed, color-separated and mobility-shift corrected, were converted to ".scf" files which were the Standard Chromatogram Format [1]. The ".scf" files were then loaded into the sequence

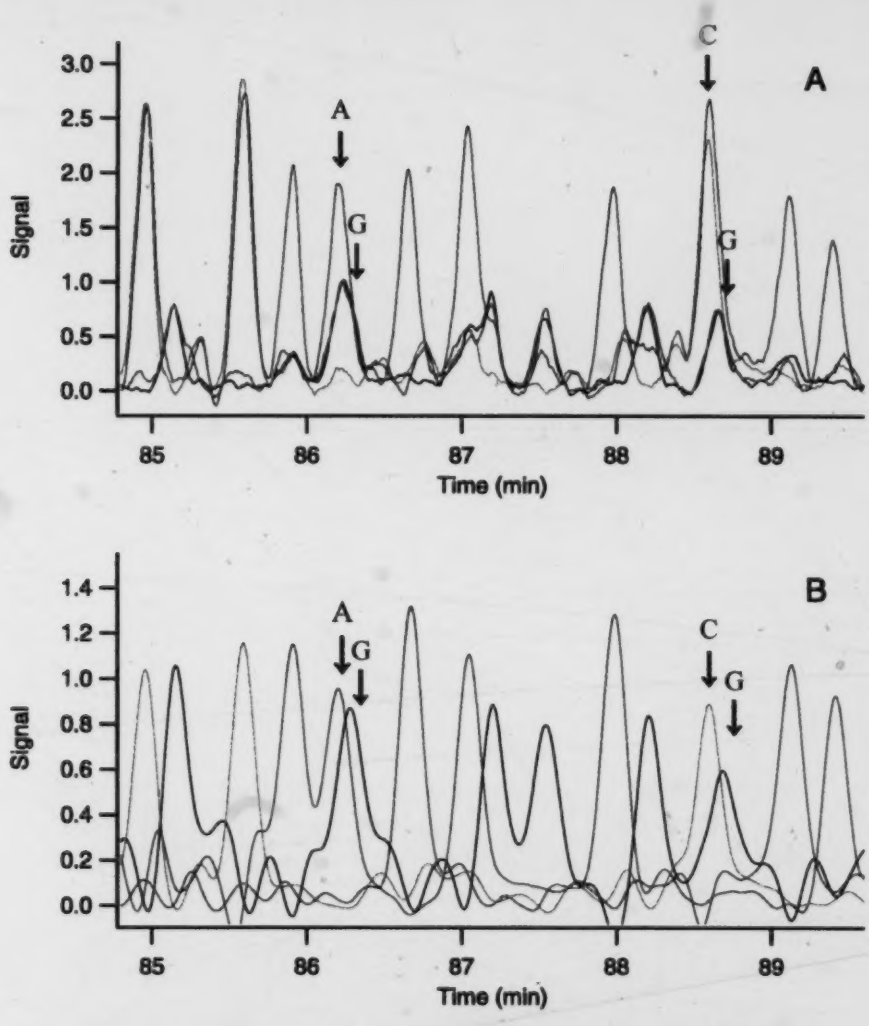


Figure 6.1. Mobility Shift of G-Terminated (TAMRA) fragments

Mobility-shifts of G-terminated fragments following A- and C-terminated fragments of *S. aureus*. (A/G, 198-199 bases after the primer and C/G, 206-207 bases after forward sequencing primer.) Figure A: raw data and Figure B: color-separated and mobility-shift corrected data.

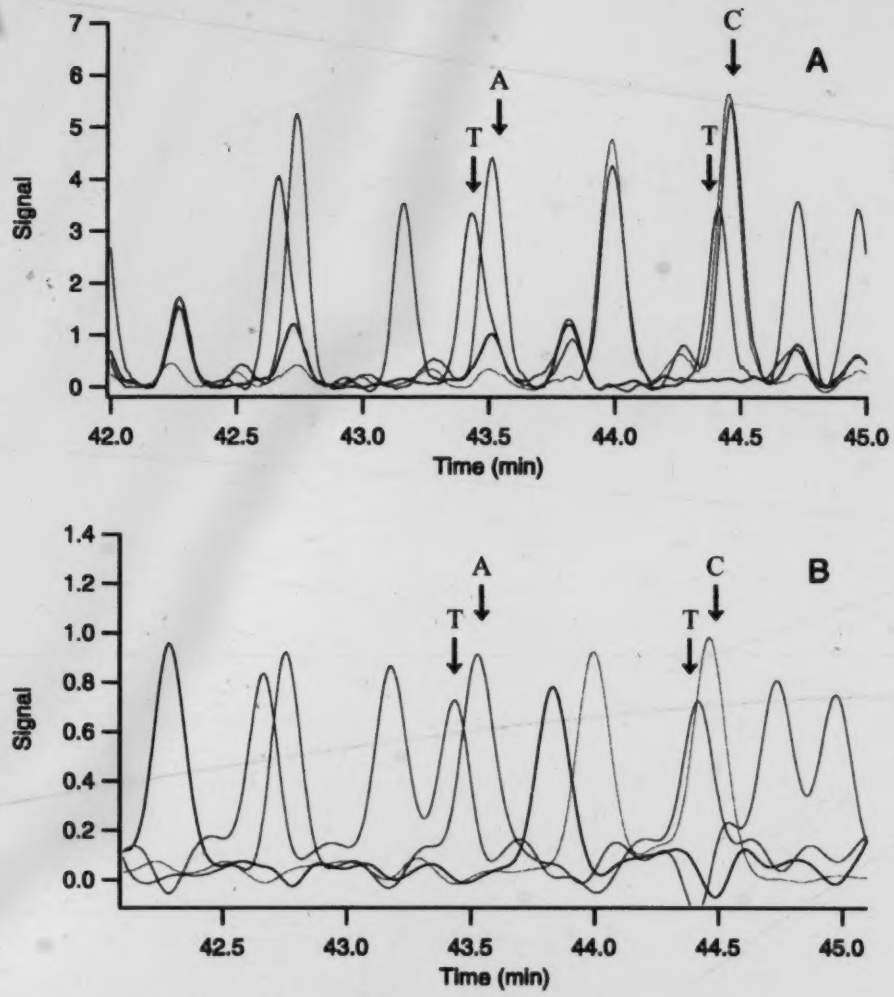


Figure 6.2. Mobility-Shift of T-Terminated (ROX) fragments
Mobility-shifts of T-terminated fragments preceding A- and C-terminated fragments of *S. aureus*. (T/A, 44-45 bases after the primer and T/C, 48-49 bases after forward sequencing primer.) Figure A: raw data and Figure B: color-separated and mobility-shift corrected data.

assembly program Sequencher V3.0 (Gene Codes Corp., Ann Arbour, MI). The ".dat" files, which were the raw data files, were converted to ".scf" files and viewed in Sequencher.

The ".scf" files were edited manually without looking at the consensus sequence. The edits were highlighted in Sequencher and the program also displayed an unalterable copy of the original unedited data. Differences in the unedited sequence and consensus sequence were called "before-editing" errors. Differences in edited data and the consensus sequences were called "after-editing" errors. At a certain point in a sequencing run, the data could not be read by the editor because the signal was too low, the resolution too poor, the color-separation was poor or the mobility-shift was too confusing. At this point, where the file were unreadable, the errors were called miscalls. Realistically, the type of errors located past 450 bases were not reliably determined; errors were probably a combination of insertions, deletions and miscalls.

In the forward-sequencing reactions, bases located after the sequencing primer, the multiple cloning site of pUC19, to the end of H279 primer sequence (totaling 32 bases) were not evaluated in this study. In the reverse-sequencing reactions, bases located after the reverse primer, the multiple cloning site of pUC19, to the end of the H280 primer sequence (73 bases) were similarly ignored.

6.2.3. Error Analysis

6.2.3.1. Data Categories

The analysis was divided into several categories:

I. Forward-primed individual sequence,

II. Reverse-primed individual sequence,

III. All forward-primed sequences, and

IV. All reverse-primed sequences.

Category I examined errors of forward -primed sequencing reactions of *S. aureus*, *S. epidermidis*, *S. haemolyticus*, *S. lugdunensis*, *S. saprophyticus*, and *S. schleiferi* sequences individually. Category II was similar to Category I except the reverse-primed sequencing reactions were analyzed. Category III looked more broadly at the errors in all

the forward-sequencing reaction runs for the 6 *Staphylococcus* sequences. Category IV was similar to Category III except all the reverse-sequencing reactions were analyzed. With the sequences having around 85% similarity, I felt that I could compare all the sequences together.

6.2.3.2. Accuracy *vs.* Base Position

A scatter plot of accuracy versus base position was determined by dividing the number of correct base-calls by the total number of bases called at that base position. The data in Categories III and IV were plotted as the percentage of correct base calls versus base position after the sequencing primer.

6.2.3.3. Distribution of Errors *vs.* 50-Base Windows

The data from Categories I to IV were examined for their distribution of errors in 50-base windows after the sequencing primer (e.g. 1-50 bases, 51-100 bases, etc.). A tally was made of the number of errors (deletion, insertion, miscall, no-call and overall) in a specified window and divided by the total number of bases called within that section. The data was plotted as the average number of errors per 50 bases.

6.2.3.4. Read Lengths

The read lengths of all the runs used were determined by calculating the overall accuracy after editing.

$$\text{Overall Accuracy} = \frac{\text{base position - cumulative errors to specified base position}}{\text{base position}} \times 100\%$$

The read lengths for each run were determined at overall accuracies of greater than 99%, 97% and 95%.

6.3. RESULTS AND DISCUSSION

6.3.1. Accuracy *vs.* Base Position

A plot of accuracy for each base position for all forward and reverse primed sequences is shown in Figures 6.3 and 6.4. It is apparent that editing improved the

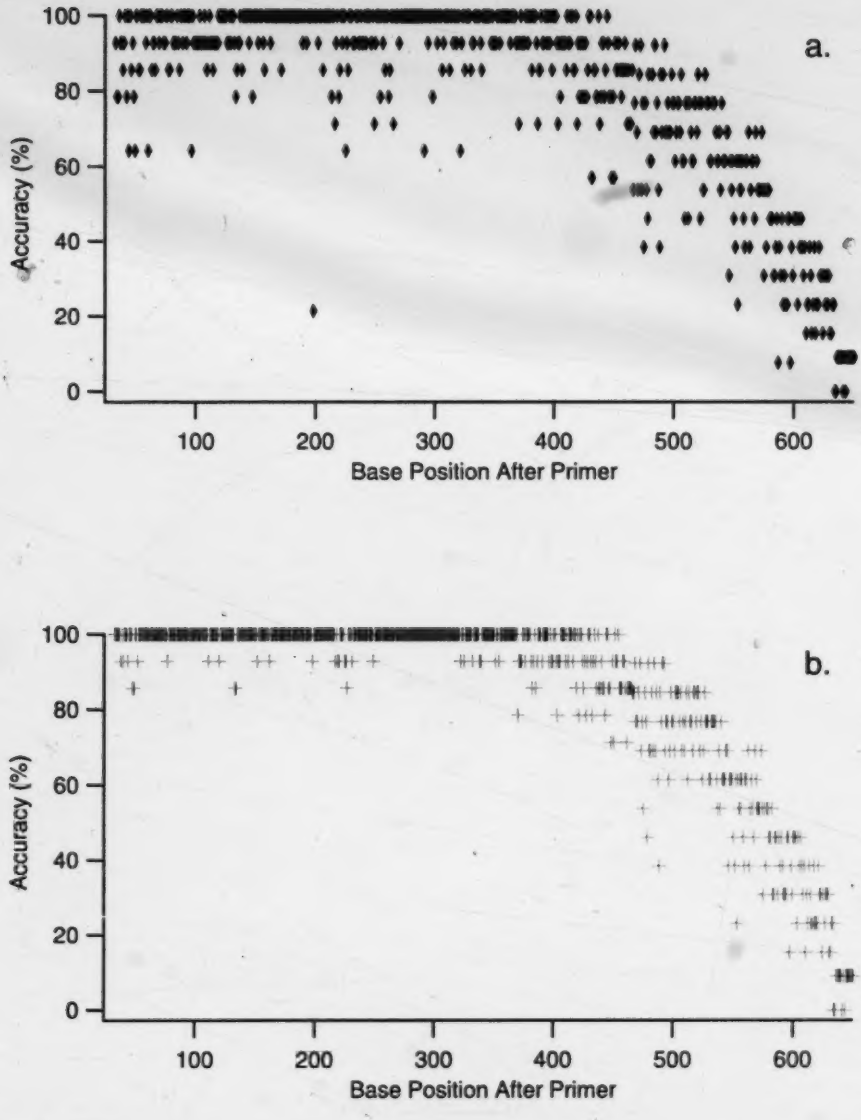


Figure 6.3. Accuracy *vs.* Base Position of All Forward-Primed Reactions.
Figure a: before-editing. Figure b: after-editing.

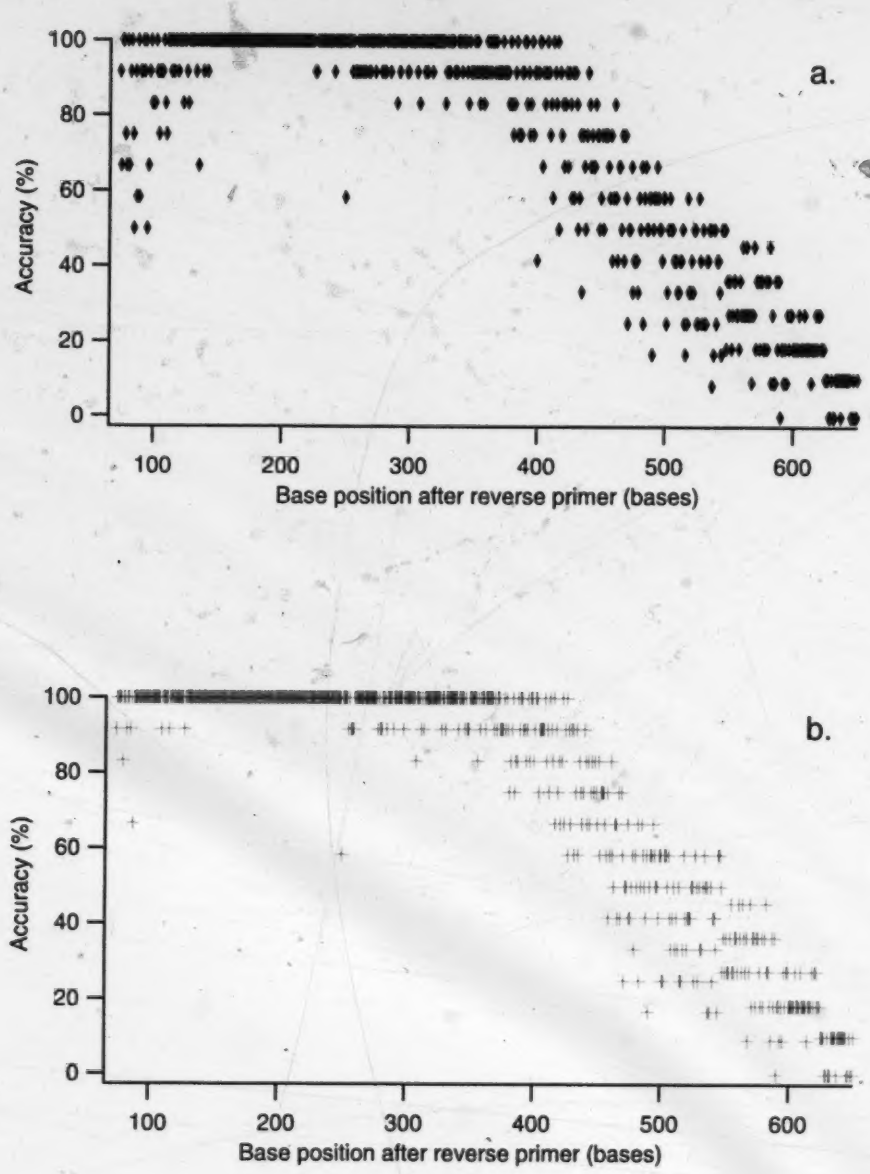


Figure 6.4. Accuracy *vs.* Base Position of All Reverse-Primed Reactions.
Figure a: before-editing. Figure b: after-editing.

accuracy for fragments shorter than 500 bases. The accuracy tended to decrease with longer fragments. This result was not surprising because the signal-to-noise and gel resolution degrades with fragment length. The large signals are expected for early migrating analytes with larger electrophoretic velocities [2]. There also may be a small sample bias in electrokinetic injections for fragments with a faster mobility [3]. This bias would cause fewer longer fragments injected and result in a decrease in signal levels over the run.

6.3.2. Distribution of Errors vs. 50-Base Windows

The distribution of errors versus 50-base windows was determined for each of the *Staphylococcus* sequences. The distribution of the overall errors, the sum of deletion, insertion, miscall and no-call errors, for each *Staphylococci* sequence (Category I and II) are plotted in Figures 6.5 and 6.6. The errors were calculated using 2 runs for each sequence, except *S. aureus* and *S. epidermidis* forward-primed sequences, where 3 runs were used.

The errors were high in the beginning of the sequences to 100-150 nucleotides after the sequencing primer, decreased in the middle and then increased dramatically at 500 bases for forward-primed reactions and 400 bases for the reverse-primed reactions. Editing improved the accuracy until late in the sequence where it was very difficult to edit the data at all.

For both *S. saprophyticus* forward-primed (Figure 6.5.e) and reverse-primed (Figure 6.6.e) reaction plots there were jumps in error rate at 201-250 base and 251-300 base windows respectively. These peaks reflected the stretch sample artifact from one of the two runs used to calculate errors. These sequences were run with a Plexiglas heater at 40°C. *S. schleiferi* forward-primed reaction (Figure 6.5.f) did not have the dramatic rise in errors at the end as observed in other *Staphylococcus* sequencing runs. One run was stopped at 466 bases and the other run was exceptionally good and errors were less than 3%, before editing, up to 670 bases after the primer.

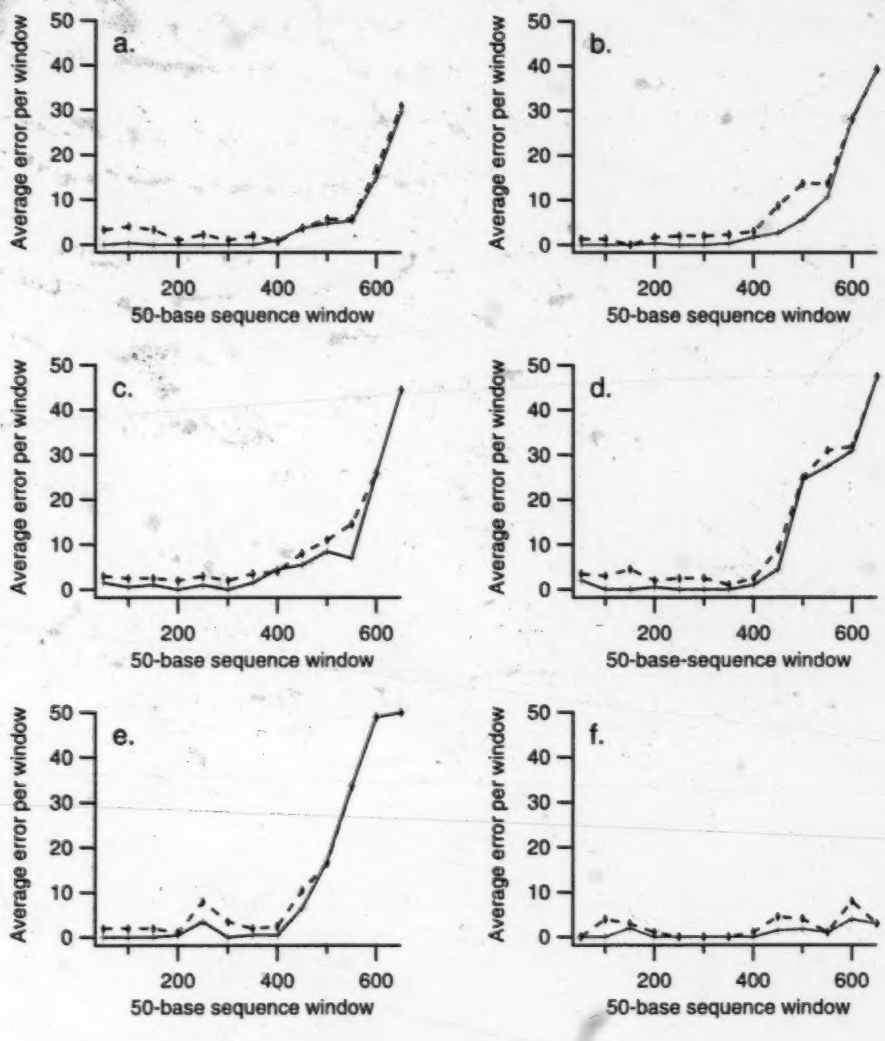


Figure 6.5. Average Error per 50-Base Window for Forward-Primed Reactions of Six Staphylococcus Species.
Figure a: *S. aureus*, b: *S. epidermidis*, c: *S. haemolyticus*, d: *S. lugdunensis*, e: *S. saprophyticus*, and f: *S. schleiferi*.

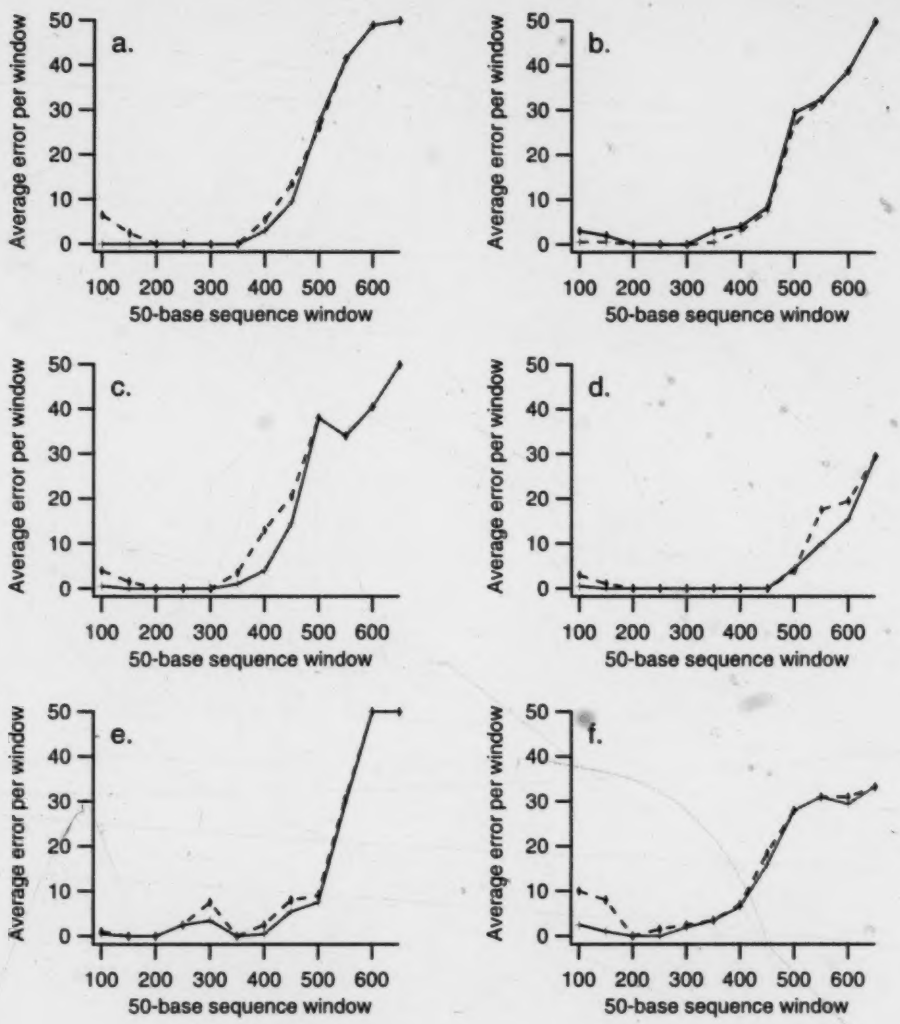


Figure 6.6. Average Error per 50-Base Window for Reverse-Primed Reactions of Six Staphylococcus Species.

Figure a: *S. aureus*, b: *S. epidermidis*, c: *S. haemolyticus*, d: *S. lugdunensis*, e: *S. saprophyticus*, and f: *S. schleiferi*.

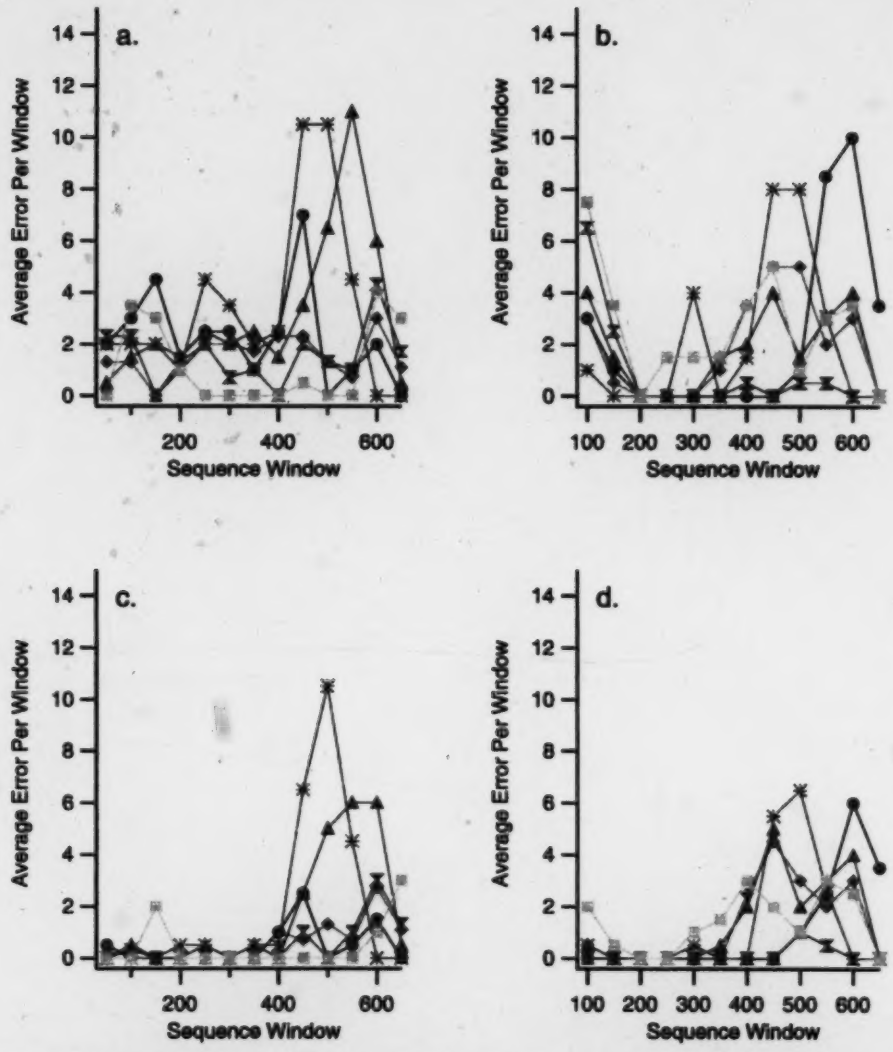


Figure 6.7. Deletion Errors per 50-Base Window for All Sequencing Reactions.
a: Forward-primed reactions, before-editing. b: Reverse-primed reactions, before-editing.
c: Forward-primed reactions, after-editing. d: Reverse-primed reactions, after-editing.

- *S. aureus*
- *S. epidermidis*
- *S. haemolyticus*
- *S. lugdunensis*
- *S. saprophyticus*
- *S. schleiferi*

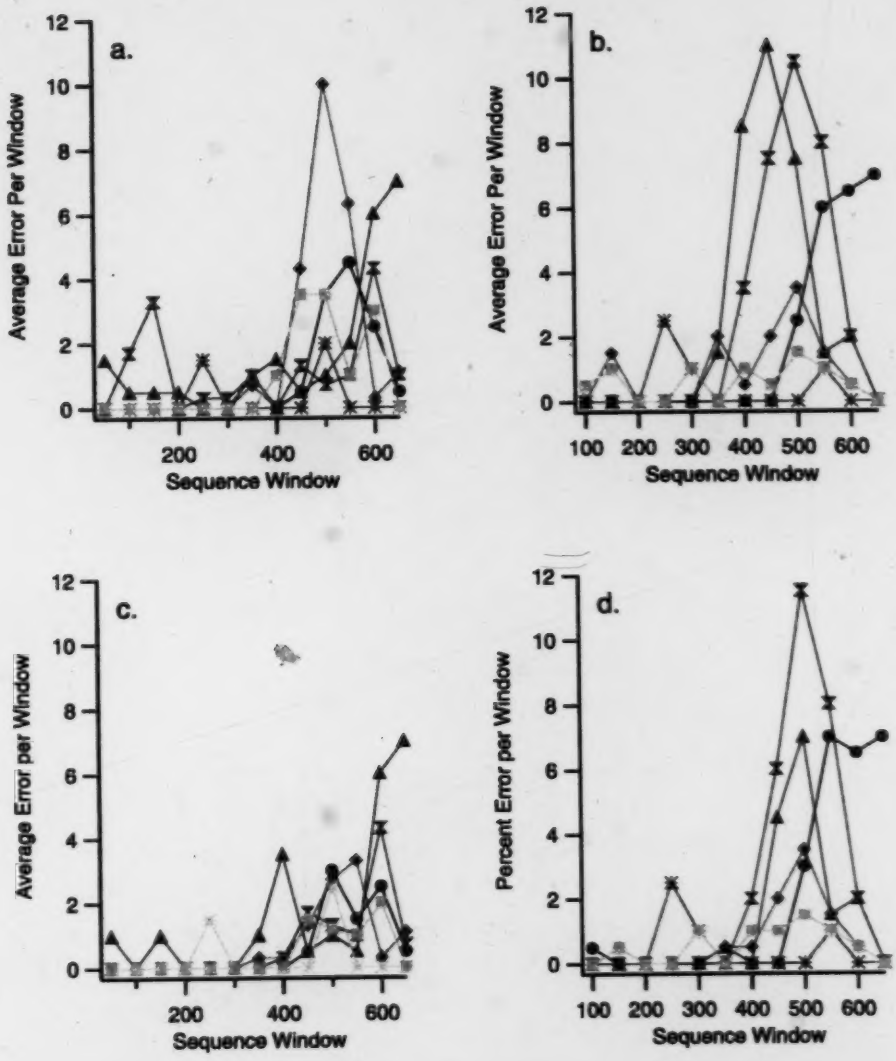


Figure 6.8. Insertion Errors per 50-Base Window for All Sequencing Reactions.
a: Forward-primed reactions, before-editing. b: Reverse-primed reactions, before-editing.
c: Forward-primed reactions, after-editing. d: Reverse-primed reactions, after-editing.

- *S. aureus*
- *S. epidermidis*
- *S. haemolyticus*
- *S. lugdunensis*
- *S. saprophyticus*
- *S. schleiferi*

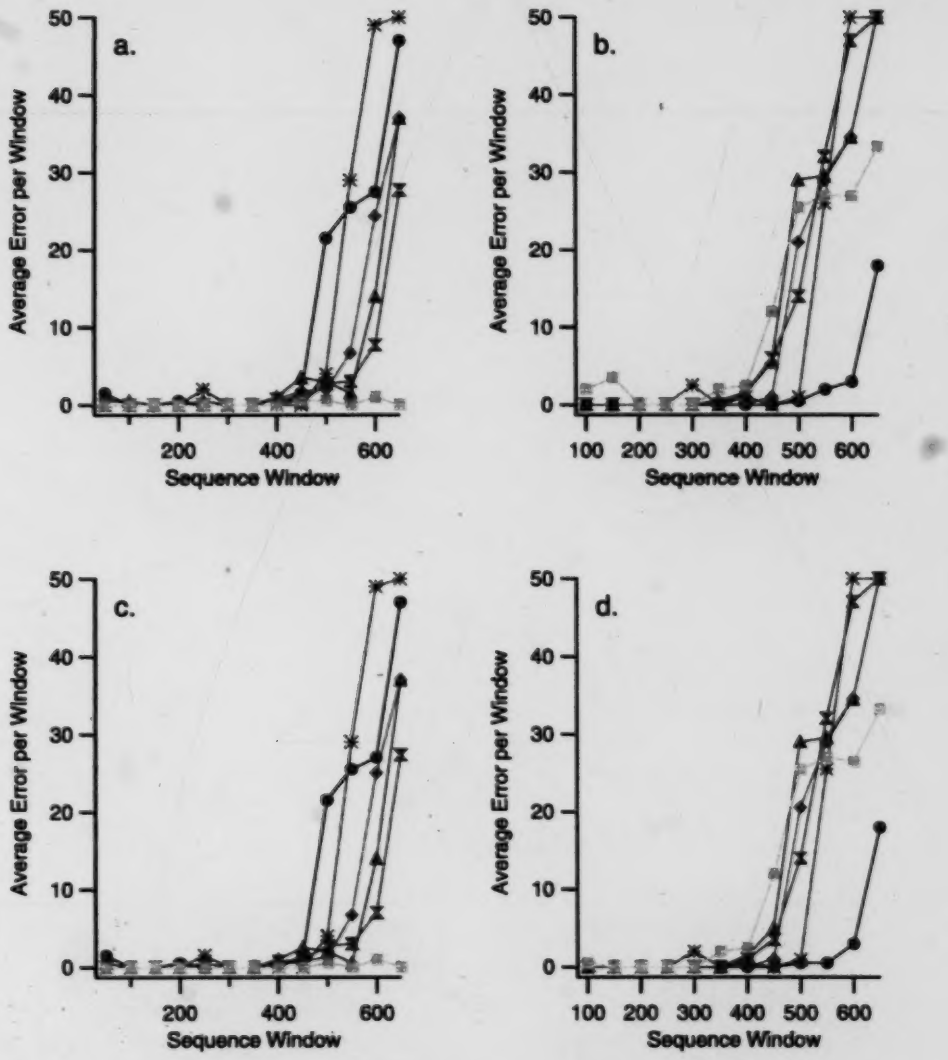


Figure 6.9. Miscall Errors per 50-Base Window for All Sequencing Reactions.
a: Forward-primed reactions, before-editing. b: Reverse-primed reactions, before-editing. c: Forward-primed reactions, after-editing. d: Reverse-primed reactions, after-editing.

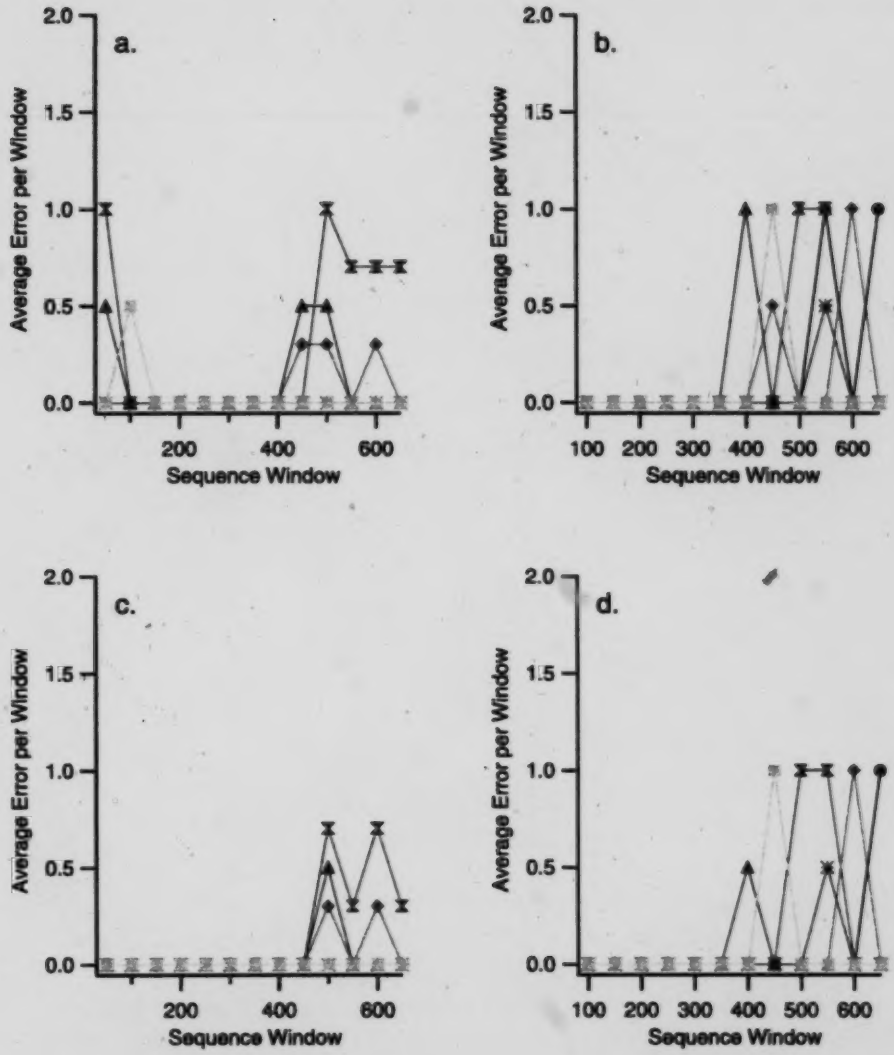


Figure 6.10. No-Call Errors per 50-Base Window for All Sequencing Reactions.
a: Forward-primed reactions, before-editing. b: Reverse-primed reactions, before-editing. c: Forward-primed reactions, after-editing. d: Reverse-primed reactions, after-editing.

- *S. aureus*
- *S. epidermidis*
- *S. haemolyticus*
- *S. lugdunensis*
- *S. saprophyticus*
- *S. schleiferi*

The breakdown of the sequence errors, deletion, insertion, miscall and no-call errors, for *Staphylococcus* sequences (Category I and II) are graphically shown in Figures 6.7 to 6.10.

The breakdown of errors for all the forward-primed reactions (Category III) and all the reverse-primed reactions (Category IV) are listed in Table 6.1. Figures 6.11 and 6.12 graphically show the overall error distribution between 0-500 bases. The average difference between edited and unedited runs was about 4% over the entire run. Editing was helpful in first 100-150 bases and was less effective for later bases.

The largest type of errors in the sequences were deletions (Figure 6.11 and 6.12). BASS had difficulty calling triplets, quartets, quintets, etc. This deletion error seemed to arise from color-separation problems. Figure 6.13, depicts a portion of reverse-primed reaction for *S. schleiferi* in capillary 13 from file dr970527-1. It shows a quintet of Ts, 398 to 402 bases after the primer. In the raw data file, the heights of the peaks were relatively even. However, after base-calling, the middle base disappeared; BASS failed to call base 400.

The biggest cause of deletion errors were attributed to dye-labeled primers creating mobility-shifted data. Deletion errors were common at the beginning of the sequences, up to 150 bases after the primer. Other groups performing accuracy studies on ABI 370 series slab gel sequencers have also found that the first 100 bases in dye-primer samples have more errors compared to dye-terminators [4-6]. Naeve *et al.* attribute the source of these errors, however, to miscalls (9% error at 0-50 base window to 6% error at 51-100 base window) due to primer-peak interference [4]. They found that editing data in this region does not improve accuracy. McDonald *et al.* found that they obtain "uneven spacing" in the first 100 bases with deletions and no-calls comprising the majority of errors [5]. Koop *et al.* found high error rates in the first 10-20 bases when sequencing with dye-labeled primers [6].

There was a sequence specific trend for some of the deletion errors caused by the mobility shift. In Table 6.2, the most common deletion sequences for the all forward- and reverse-primed reactions are listed. The capital letter represents the base missed by base-caller.

Table 6.1.a. Error Distribution of Forward-Primed Reactions per 50-Base Window

Window	Del be	Del ae	Ins be	Ins ae	Mis be	Mis ae	NoC be	NoC ae	Total be	Total ae
33-50	7.9%	0.4%	1.2%	0.8%	1.6%	1.6%	1.6%	0.0%	12.3%	2.8%
51-100	4.4%	0.3%	0.9%	0.0%	0.1%	0.0%	0.1%	0.0%	5.6%	0.3%
101-150	3.3%	0.6%	1.6%	0.3%	0.0%	0.0%	0.0%	0.0%	4.9%	0.9%
151-200	2.4%	0.1%	0.1%	0.0%	0.3%	0.3%	0.0%	0.0%	2.9%	0.4%
201-250	4.4%	0.3%	0.6%	0.4%	0.7%	0.6%	0.0%	0.0%	5.7%	1.3%
251-300	3.4%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%	0.0%
301-350	2.7%	0.3%	1.0%	0.4%	0.0%	0.0%	0.0%	0.0%	3.7%	0.7%
351-400	2.9%	0.9%	0.7%	1.3%	0.9%	0.7%	0.0%	0.0%	4.4%	2.9%
401-450	8.0%	4.0%	3.7%	2.0%	2.4%	1.9%	0.3%	0.0%	14.4%	7.9%
451-500	6.1%	5.1%	7.4%	3.9%	10.5%	9.9%	0.7%	0.6%	24.8%	19.6%
501-550	5.8%	4.2%	5.5%	2.8%	21.7%	21.4%	0.3%	0.2%	33.4%	28.5%
551-600	6.5%	5.1%	5.2%	5.1%	42.8%	42.6%	0.5%	0.5%	54.9%	53.2%
601-650	1.9%	1.7%	3.3%	3.1%	71.0%	70.9%	0.3%	0.2%	76.5%	75.9%

Table 6.1.b. Error Distribution of Reverse-Primed Reactions per 50-Base Window

Window	Del be	Del ae	Ins be	Ins ae	Mis be	Mis ae	NoC be	NoC ae	Total be	Total ae
74-100	15.4%	2.2%	0.3%	0.3%	1.2%	0.3%	0.0%	0.0%	17.0%	2.8%
101-150	3.0%	0.2%	0.8%	0.3%	1.2%	0.0%	0.0%	0.0%	5.0%	0.5%
151-200	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
201-250	0.5%	0.0%	0.8%	0.8%	0.0%	0.0%	0.0%	0.0%	1.3%	0.8%
251-300	1.8%	0.5%	0.7%	0.7%	0.8%	0.7%	0.0%	0.0%	3.3%	1.8%
301-350	1.3%	0.7%	1.2%	0.3%	0.8%	0.7%	0.0%	0.0%	3.3%	1.7%
351-400	3.7%	2.5%	4.5%	1.2%	2.2%	1.8%	0.3%	0.2%	10.7%	5.7%
401-450	7.3%	5.7%	7.0%	4.5%	8.2%	7.2%	0.5%	0.3%	23.0%	17.7%
451-500	5.7%	4.8%	8.5%	8.8%	30.3%	30.2%	0.3%	0.3%	44.8%	44.2%
501-550	6.7%	4.5%	6.4%	6.7%	48.5%	47.8%	0.8%	0.5%	62.4%	59.5%
551-600	7.5%	5.6%	4.2%	4.2%	62.2%	62.0%	0.4%	0.4%	74.2%	72.2%
601-650	1.3%	1.3%	2.7%	2.7%	83.0%	83.0%	0.4%	0.4%	87.4%	87.4%

Deletion (Del); Insertion (Ins); Miscall (Mis); No-call (NoC); before-editing (be); after-editing (ae).

Table 6.2. Common Deletion Error Sequences

Forward-Primed Sequence	Frequency (out of 8472 bases)	Reverse-Primed Sequence	Frequency (out of 6797 bases)
aAg	63	tTc	35
aAa	32	tTt	20
cAg	21	cCa	18
tTa	16	tAc	18
aAt	13	aTc	14
tAg	12	gCa	14

Table 6.3. Frequency of AAG and TTG Sequences in *Staphylococcus* Sequences

	Frequency of AAG	Frequency of TTC
Forward-Primed Sequences	124	49
Reverse-Primed Sequences	30	112

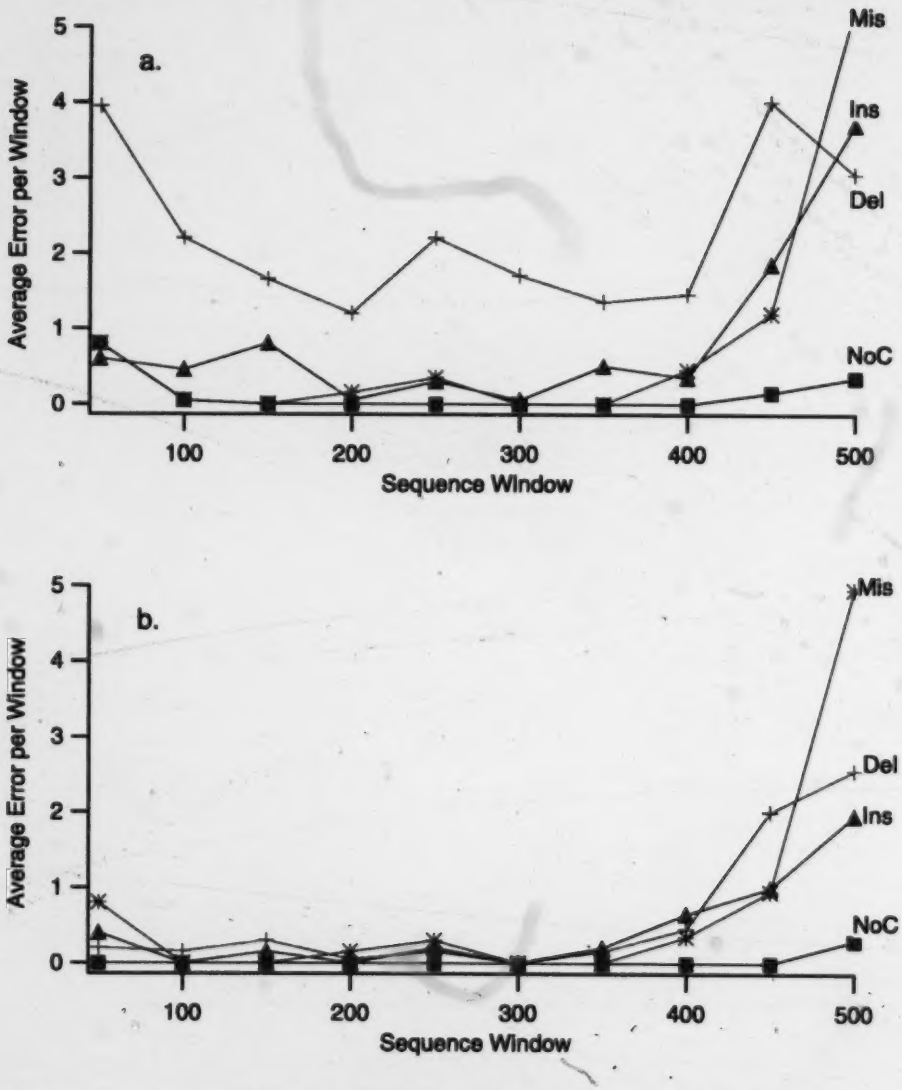


Figure 6.11. Errors of All Forward-Primed Reactions per 50-Base Window.
a: before-editing. b: after-editing.

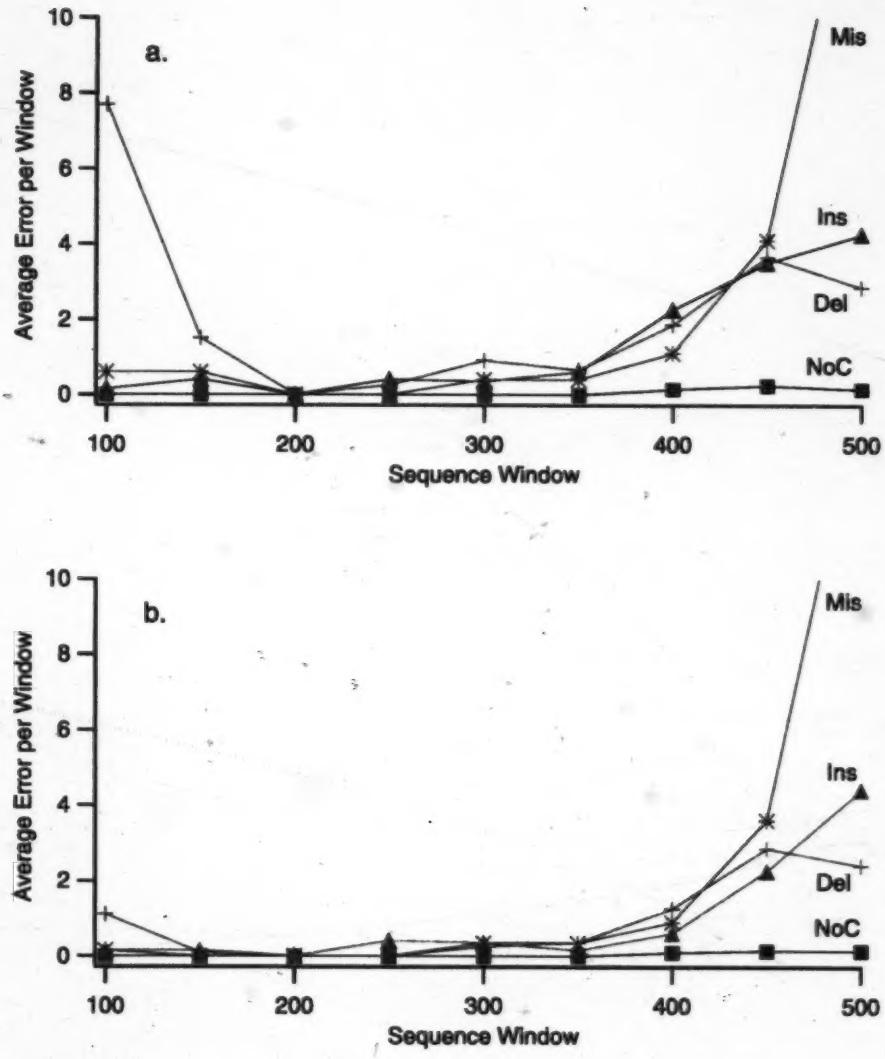


Figure 6.12. Errors of All Reverse-Primed Reactions per 50-Base Window.
a: before-editing. b: after-editing.

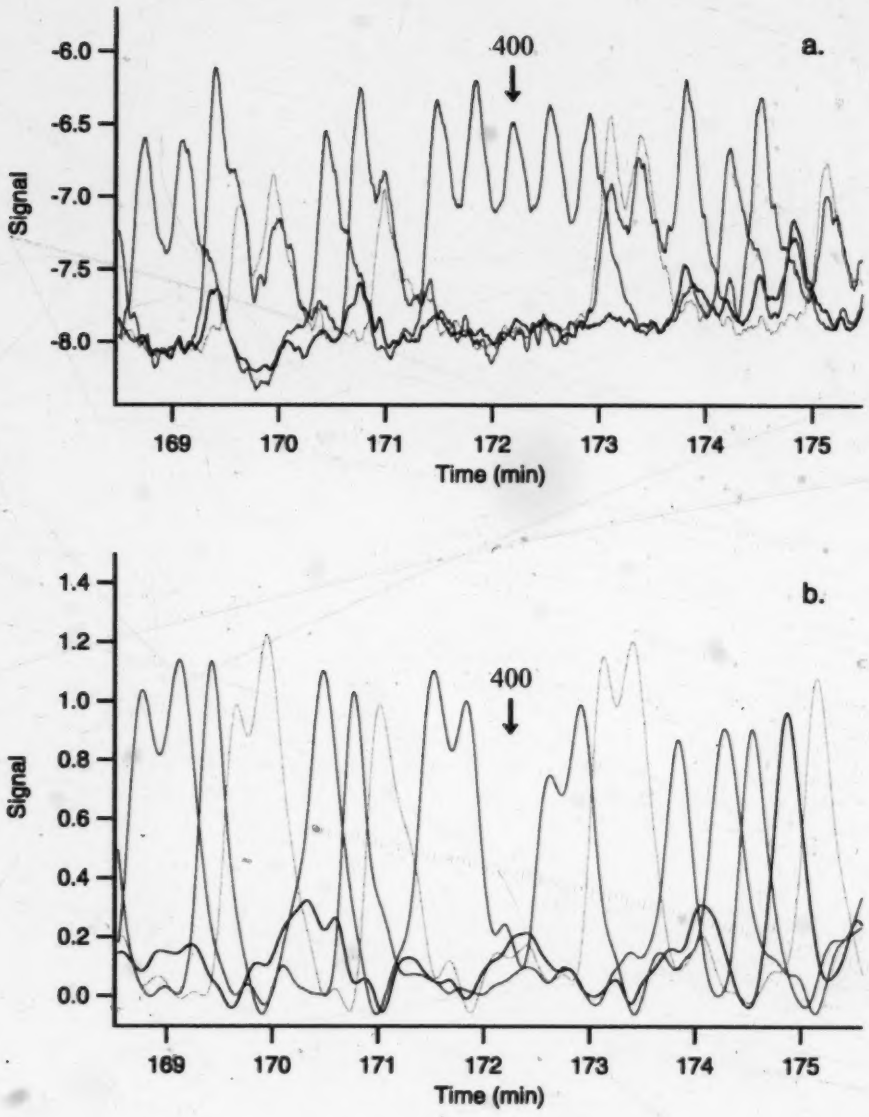


Figure 6.13. Color-Separation of Multiplets.
A quintet of Ts, 398 to 402 bases after the reverse sequencing primer in *S. schleiferi*. After data processing, base 400 disappears. a: raw data and b: color-separated data.

The most common error in the forward-primed sequencing reactions was associated with the mobility shift of G following an A. G moved more quickly and tended to overlap with the shorter A fragment. This sequence specific error is shown in Figure 6.1.a, where there is a drop in the accuracy to 21.4% at 198 bases. This point was an aAg deletion error that occurred in 11 out of 14 runs (representing all 6 *Staphylococcus* sequences). The next most common sequence specific error in the forward strand was deleting an A in a multiplet of A. The most common error in the reverse strand was the mobility shift of a T before a C. T moved more slowly through acrylamide gel, about one base slower if followed by a C.

The reverse strand had a different sequence specific deletion pattern compared to the forward strand. I attributed this variation to the sequence difference in the forward and reverse strands. The frequency of the occurrence of AAG and TTC in the forward and reverse directions in consensus sequence of the six *Staphylococcus* species is shown in Table 6.3.

Mobility shifts had a profound affect on BASS. Mobility shifts problem can be solved in either the software or the DNA fluorescent sequencing chemistry. The program can be written to take into account the A-G and T-C sequence dependence mobility shifts. Another solution would be to use either dye-terminators to label the fragments. Dye-terminator reactions can be base-called earlier in the sequence around 25 bases after the primer [4] because of the reduced mobility shift effects. The trade-off is that dye terminator reactions tend to have shorter read lengths, only reading out to 300-350 bases with higher errors on ABI sequencers [4, 6, 7]. The higher error rate is attributed to the less uniform peak heights of dye-terminators when compared to the dye-primers. Another solution could be to use other dye-labeled primers such as the Energy-Transfer Dyes [8], or BODIPY dyes which have specially modified linker arms to minimize mobility shifts [9].

Insertion errors were the second common error type, before the exponential increase in errors around 400-500 bases (Figure 6.11 and 6.12). Table 6.4 lists the frequency of insertion errors. Insertion errors usually resulted from color separation problems. Typically the weakest signals were for G-terminated fragments. Despite

Table 6.4. Frequency of Inserted Base

Inserted Base	Forward Sequence Base Range	Frequency (before edit)	Reverse Sequence Base Range	Frequency (before edit)
A	111-630	62	250-648	38
C	233-624	56	91-637	58
G	38-627	75	79-645	100
T	227-584	11	250-623	18
N	567	1	250-488	3

Table 6.5. Most Common Miscall Errors

Forward Strand Miscall	Frequency	Reverse Strand Miscall	Frequency
A for C	55	G for T	30
G for A	25	C for T	29
C for A	24	C for A	23
G for T	9	G for A	18
A for T	7	G for C	18
C for G	5	C for G	11
		A for C	9
		A for T	9

sample preparation ratios of A:C:G:T of 1:1:2:2, G peaks remained low. As a result, the color separation had difficulty calculating the color matrix for G. Usually BASS tended to overcall G peaks; G insertion errors were the largest.

Miscall errors also arose from color separation problems. Table 6.5 indicates the most common miscalls for all the forward-primed and all reverse-primed sequences. X for Y indicated that a base was called X when it was really Y.

6.3.3. Read Lengths

The read lengths of forward, reverse and overall runs are presented in Tables 6.6, 6.7 and 6.8. The accuracy was calculated after-editing. It was interesting to note that the reverse reactions give consistently shorter read lengths than the forward-primed reactions. At the specified accuracy, the reverse sequences were approximately 60 bases shorter. This result was puzzling because there should be no difference in the forward- or reverse-primed sequencing reactions.

ABI reports read lengths of 450 bases at 98% accuracy for the 373 slab gel sequencer [10], 600 bases at >98.5% accuracy in 3 h for the 377 slab gel sequencer [11] and 600 bases at >98.5% accuracy in 165 min for the ABI 310 capillary sequencer [12]. The reported accuracies omit no-call errors, which may contribute up to 2% of the total errors. The average read lengths of the 16-capillary sequencer were about 500 bases at >97% accuracy in about 3 h.

The read length of the 16-capillary sequencer was disappointing because 1000 base reads of single-stranded M13mp18 were demonstrated on capillaries at elevated temperatures and linear polyacrylamide gels [13-15]. There were several possible reasons why the read lengths were shorter for the HSP60 clones. First, the quality of gel coating to the capillary wall plays an important role in successful capillary sequencing. Gels prepared for this thesis varied from batch-to-batch and even varied within a batch. The within-batch variation was most likely due to variations in gel coating in the capillaries.

Second, the sequencing template also plays a significant role in the quality of the sequencing data. Sequencing single-stranded templates such as M13 usually generated the best sequencing data. However single-stranded templates are laborious to prepare and

Table 6.6. Read Length at >99% Accuracy

	Forward Sequences	Reverse Sequences	Forward and Reverse
Average (bases)	463	409	437
Median (bases)	480	389	452
Maximum (bases)	566	537	566
Minimum (bases)	217	250	217
Stdev (bases)	99	81	93
n	12	11	23

Table 6.7. Read Length at >97% Accuracy.

	Forward Sequences	Reverse Sequences	Forward and Reverse
Average (bases)	522	444	486
Median (bases)	517	456	489
Maximum (bases)	670	593	670
Minimum (bases)	372	255	255
Stdev (bases)	91	101	102
n	14	12	26

Table 6.8. Read Length at >95% Accuracy

	Forward Sequences	Reverse Sequences	Forward and Reverse
Average (bases)	553	491	520
Median (bases)	544	468	525
Maximum (bases)	650	625	650
Minimum (bases)	426	349	349
Stdev (bases)	75	88	86
n	10	11	21

most laboratories prefer to sequence double-stranded templates. In a recent survey of core DNA sequencing facilities, 75% of the templates sequenced were double-stranded plasmids, 15% were PCR products and 10% were single-stranded templates [16]. There remains a trade-off between convenience, time and best quality sequence data. PCR products were sequenced since they would most likely be the templates sequenced for the HSP60 diagnostic.

Third, the sensitivity of the sequencing instrument enables the detection of longer fragments. The 16-capillary sequencer was not as sensitive as a single-capillary instrument. The 16-capillary instrument had a mass limit of detection of ~10 000 molecules, and a concentration detection limit of 10^{-12} M fluorescein. The single-capillary instruments have mass detection limits of hundreds of molecules with microscope objectives with higher numerical aperture of 0.6 [17].

6.4. CONCLUSIONS

The capillary sequencing accuracy was comparable to the commercially available ABI sequencers. I reported a capillary DNA sequencing accuracy of >97% for about 500 base read lengths using dye-primer chemistry and double-stranded PCR templates.

The base-calling software required manual editing of data, which improved sequencing accuracy on average 4%. It was apparent that extensive manual editing of the data, however, was impractical and unacceptable for future sequencing projects. Manual editing required an experienced editor and as a result, editing may be subjective and vary from user to user. In the HSP60 diagnostic, the accuracy of DNA sequencing was extremely important to correctly identify the *Staphylococcus* species infecting the patient. For the six *Staphylococcus* species used in this study, the DNA sequences only differed by 15% overall.

BASS requires modification to correct for mobility shifts, base spacing and color separation when the data is less than perfect. It was not surprising that the quality of the base-calling depends on signal-to-noise and peak resolution. The parameters can be adjusted with appropriate sample preparation chemistry and choice of gel matrix.

6.5. REFERENCES

(1) Voss, K. O., CONVERT Program, University of Alberta, **1997**.

(2) Cheng, Y. F.; Wu, S.; Chen, D. Y.; Dovichi, N. J. *Analytical Chemistry* **1990,** *62*, 496-503.

(3) Huang, X.; Gordon, M. J.; Zare, R. N. *Analytical Chemistry* **1988,** *60*, 375-377.

(4) Naeve, C. W.; Buck, G. A.; Niece, R. L.; Pon, R. T.; Robertson, M.; Smith, A. J. *BioTechniques* **1995,** *19*, 448-452.

(5) McDonald, L. A.; Kelley, J. M.; Brandon, R. C.; Adams, M. D. *BioTechniques* **1995,** *19*, 464-470.

(6) Koop, B. F.; Rowan, L.; Chen, W.-Q.; Deshpande, P.; Lee, H.; Hood, L. *BioTechniques* **1993,** *14*, 442-447.

(7) Pogue, R. P.; Cook, M. E.; Livingstone, L. R.; Hunt III, S. W. *BioTechniques* **1993,** *15*, 377-379.

(8) Ju, J.; Ruan, C.; Fuller, C. W.; Glazer, A. N.; Mathies, R. A. *Proc. Natl. Acad. Sci (USA)* **1995,** *92*, 4347-4351.

(9) Metzker, M. L.; Gibbs, R. A., *Automation in Mapping and DNA Sequencing Conference*, Lawrence Berkeley National Laboratory **1995**; 66.

(10) ABI PRISM 373 DNA Sequencer, Perkin-Elmer/Applied Biosystems, **1995**.

(11) ABI PRISM 377 DNA Sequencer, Perkin-Elmer/Applied Biosystems, **1997**.

(12) ABI PRISM 310 Genetic Analyzer, Perkin-Elmer/Applied Biosystems, **1997**.

(13) Zhang, J.-Z.; Fang, Y.; Hou, J. Y.; Ren, H. J.; Jiang, R.; Roos, P.; Dovichi, N. J. *Analytical Chemistry* **1995,** *67*, 4589-4593.

(14) Kleparnik, K.; Foret, F.; Berka, J.; Goetzinger, W.; Miller, A. W.; Karger, B. L. *Electrophoresis* **1996,** *17*, 1860-1866.

(15) Carrilho, E.; Ruiz-Martinez, M. C.; Berka, J.; Smirnov, I.; Goetzinger, W.; Miller, A. W.; Brady, D.; Karger, B. L. *Analytical Chemistry* **1996,** *68*, 3305-3313.

(16) Pon, R. T.; Buck, G. A.; Niece, R. L.; Robertson, M.; Smith, A. J.; Spicer, E. *BioTechniques* **1994,** *17*, 526-534.

(17) Swerdlow, H.; Zhang, J. Z.; Chen, D. Y.; Harke, H. R.; Grey, R.; Wu, S.; Dovichi, N. J. *Analytical Chemistry* **1991,** *63*, 2835-2841.

Chapter 7: DNA Sequencing Artifacts

7.1. INTRODUCTION

The majority of the genome sequencing groups use single-stranded (ss) templates for sequencing DNA [1-3]. These templates give the best quality sequence data, but are laborious to prepare. A survey of the core DNA sequencing facilities, which lack the resources of the genome projects, found that double-stranded (ds) samples were the most preferred template: 75% of the samples were ds templates, 15% were PCR samples and 10% were ss templates [4]. PCR samples as sequencing templates was reported to be a fast growing trend [5-7].

This chapter describes the capillary DNA sequencing artifacts observed while sequencing wild type (wt) M13 on the first generation instrument and *Staphylococcus* samples and the second generation instrument. Dirty PCR sequencing templates artifacts were eliminated by cleaning the PCR templates. Secondary DNA structure artifacts were eliminated using higher temperatures during electrophoresis.

7.2. EXPERIMENTAL

7.2.1. DNA Sequencing Samples

AB PCR Reaction

Fragments of wild type M13 or the Heat Shock Protein 60-kDa (HSP60) gene from *Staphylococcus* were cloned into pUC19 plasmid. PCR templates for sequencing were generated by amplifying across the pUC19 multiple cloning site. The primer sequences were 5'-biotin-GCTTCCGGCTCGTATGTTGTGTG-3' for the A1 primer and 5'-biotin-AAAGGGGGATGTGCTGCAAGGCG-3' for the B2 primer (University of Calgary Core DNA Lab, Calgary, Alberta). The A1 priming site was located at position 509 to 531 bp of pUC19. The B2 priming site was located at 315 to 337 bp.

Wild-Type M13 Sequencing

A single white colony was selected with a pipet tip and dipped into 50 µl of water. The bacterial cells were lysed by boiling the water and the cellular debris was pelleted by a quick spin in the microcentrifuge at 13 000 rpm. 5 µL of the supernatant was added to 38 ng of primer A1 and B2 (5 pmol), 10 nmoles dNTPs, 75 nmoles $MgCl_2$, 5 µl of

10×PCR reaction buffer and 2.5 U Taq (Gibco BRL, Gaithersburg, MD). The final volume of the reaction was 50 µL. The thermal cycling conditions were: 30 s at 96°C, 25 cycles of 30 s at 94°C, 30 s at 55°C and 1 min at 72°C, and last cycle of 2 min at 72°C. 10 µL of the AB PCR products were run on a gel to verify the size. Positive clones (pUC19 with wt M13 inserted) produced a 700 to 1700 bp PCR product. AB PCR reaction was precipitated in 2.5 volumes of 98% ethanol and resuspended in 1×TE (10 mM Tris-HCl and 1 mM EDTA, pH=8.0).

The cycle sequencing reactions were prepared using Amersham RPN 2536 kit with ThermoSequenase DNA polymerase (Amersham Life Sciences) and dye-labeled primers (Kit 401487, PE Applied Biosystems, Foster City, CA). The recipe in Table 7.1 was used to prepare a single cycle-sequencing reaction. The four terminations were prepared in four separate tubes. The reagent mix contained the dNTPs, the enzyme, buffer and the appropriate ddNTP. Approximately 275 ng total of double stranded PCR template was required for sequencing reactions for the cycle sequencing recipe shown in Table 7.1.

Table 7.1. Wild-Type M13 Cycle Sequencing Sample

	A-termination	C-termination	G-termination	T-termination
Reaction Ratio A:C:G:T	1	1	1.5	2
Template	50 ng	50 ng	75 ng	100 ng
M13 Forward or Reverse Dye-Primer	0.8 pmol (JOE)	0.8 pmol (FAM)	1.2 pmol (TAMRA)	1.6 pmol (ROX)
Reagent Mix	4 µL (ddATP)	4 µL (ddCTP)	6 µL (ddGTP)	8 µL (ddTTP)
Final Volume	16 µL	16 µL	24 µL	32 µL

Cycle sequencing was performed for 30 cycles of 30 s at 95°C and 30 s at 55°C. The four tubes were combined in a Microcon 30 tube (Amicon, Beverly, MA). The tubes were spun at 11000 rpm for 4-5 min in a microcentrifuge. The membrane was washed with 200 µL of water. The tubes were spun again for 4-5 min at 11000 rpm. The tubes were inverted into a new tube and the cleaned DNA was spun out of membrane in a small volume of water at 6000 rpm for 4-5 min.

Staphylococcus Sequencing

A single white colony was selected with a pipet tip and dipped into 20 µl of water. Boiling the water lysed the bacterial cells and the cellular debris was pelleted by a quick spin in the microcentrifuge. 10 µL of the supernatant was added to 38 ng of primer A1 and B2 (5 pmol), 10 nmoles dNTPs, 75 nmoles $MgCl_2$, 5 µl of 10×PCR reaction buffer and 2.5 U Taq. The final volume of the reaction was 50 µL. The thermal cycling conditions were: 30 s at 96°C, 25 cycles of 30 s at 94°C, 30 s at 55°C and 1 min at 72°C, and last cycle of 2 min at 72°C. 4 µL of AB PCR reaction was run on agarose gel. Positive clones (pUC19 with HSP60 PCR product inserted) produced a 800 bp PCR product. The positive AB PCR products were cleaned with QIAquick PCR purification kit (Qiagen Inc, Chatsworth, CA) and resuspended in 1×TE.

The cycle sequencing reactions were prepared as above. The templates used were AB PCR products from the original 600 bp HSP60 or halved HSP60 fragments cloned into pUC19. Approximately 180 ng total of double stranded PCR template was required for sequencing reactions for the cycle sequencing recipe shown in Table 7.2.

Table 7.2. Dye-Labeled Primer Cycle-Sequencing Reaction

	A-termination	C-termination	G-termination	T-termination
Reaction ratio A:C:G:T	1	1	2	2
PCR template	30 ng	30 ng	60 ng	60 ng
M13 forward or reverse dye-primer	0.8 pmol (JOE)	0.8 pmol (FAM)	1.6 pmol (TAMRA)	1.6 pmol (ROX)
reagent mix	4 µL	4 µL	8 µL	8 µL
final volume	16 µL	16 µL	32 µL	32 µL

The thermal cycling conditions were: 30 cycles of 30 s at 95°C and 30 s at 55°C.

The 4 reaction tubes were combined and the sequencing samples were desalted and cleaned with the following protocol. 15 µL of streptavidin immobilized on 4% beaded agarose (Sigma, St. Louis, MO) was added and gently mixed for 20 minutes at room temperature in the dark. The streptavidin fixed the biotinylated AB PCR templates. The agarose beads were pelleted with a quick spin in the microcentrifuge and the

supernatant was applied to a Microcon 30 (Amicon) membrane. The tubes were spun at 11000 rpm for 4-5 min with the microcentrifuge. The membrane was washed with 200 µL of 0.1×TE (1 mM Tris-HCl and 0.1 mM EDTA, pH=8.0). The tubes were spun again for 4-5 minutes at 11000 rpm. The tubes were inverted into a new tube and the cleaned DNA spun out of membrane after 4-5 min at 6000 rpm. The sample volumes ranged from 5 to 20 µL.

7.2.2. Gel Preparation

Capillaries were 50 µm ID, 150 µm OD, and typically 40 cm long (Polymicro Technologies, Phoenix, AZ). The inner walls of the capillaries were coated using Hjerten's method to covalently bind the polyacrylamide to the walls [8]. A solution of 20 µl of γ-methacryloxypropyltrimethoxysilane (Sigma) in 1 mL of 95% ethanol was drawn through the capillaries for 1 h.

To polymerize the gel in the capillaries, the injection ends were placed in a 50-ml 3-neck flask. The acrylamide solution was 6%T in 7 M urea, 0.6 g acrylamide and 4.2 g urea diluted to 10 ml with 1×TBE buffer (89 mM Tris, 89 mM boric acid and 2.0 mM EDTA, pH=8.0). The solution was passed through a 0.22 µm filter and added to the flask. The solution was degassed with a vacuum pump for 10 minutes. The polymerization of the acrylamide was initiated under an argon blanket by adding 10 µL of 10% w/v ammonium persulfate and 5 µL of TEMED (N,N,N',N'-tetramethylethylenediamine). The solution was immediately drawn through the capillaries by vacuum, with the detector ends of the capillaries under water. The vacuum was removed and the gel was left to polymerize overnight. Capillaries were visually inspected for bubbles before insertion into the cuvette.

7.2.3. Sample Injection and Run

Sixteen gel-filled capillaries were inserted into the sheath-flow cuvette and pre-run with aligning solution of 0.4 nM JOE-labeled primer and 0.4 nM of ROX-labeled primer in 1×TBE buffer at 150 V/cm. After the alignment of the instrument was confirmed, the capillaries were heated to the desired temperature using either the

Plexiglas heater described in Section 2.2.1.2 or the Peltier heater described in Section 3.2.1.3. The aligning solution was then migrated from the capillary by electrophoresis of 1×TBE. At the desired temperature, the samples were injected onto the capillaries for 30 s at 100 V/cm. The sample vial was replaced with a vial containing polymerized gel in 1×TBE. The electrophoresis was continued at 150 V/cm.

7.3. RESULTS AND DISCUSSION

There were two general types of sample artifacts, which I have termed stretches and lumps. Stretches were characterized by a region of degraded resolution, followed by a gap without sequencing peaks, a region with poor resolution, and then improved resolution. Lumps were any fluorescent peaks that obscured the sequencing fragment peaks. Examples of these artifacts can be found in the following discussion.

Wild-Type M13 Sequencing

The wild-type M13 sequencing exercise was our first attempt to sequence large amounts of DNA. PCR templates were much easier to prepare than double-stranded plasmids, which were easier to prepare than single-stranded templates. However, we learnt very quickly that clean templates were essential for DNA sequencing.

Figure 7.1 shows the catastrophic results of preparing a DNA sequencing sample from a dirty PCR template. The sample was from file sb960616-1_cap6, run at 40°C. There was a stretch artifact around 55-60 min followed by a sharp lump around 73-75 min and broad lumps from 75 to 85 min. Then at the end there were a series of 2 min wide lumps, evenly spaced at about 5 minutes from 170 to 200 min.

The lump artifacts were only found in sequencing sample prepared from PCR templates; lumps were not observed in samples prepared from double-stranded plasmids templates (Figure 2.21). In Figure 7.1, residual RNA binding to the template and priming DNA sequencing probably caused the lumps in the middle of the electropherogram. Precipitating the DNA in ethanol before DNA sequencing cleaned the AB PCR products. RNA would also be precipitated at this step, remaining with the DNA template. RNA

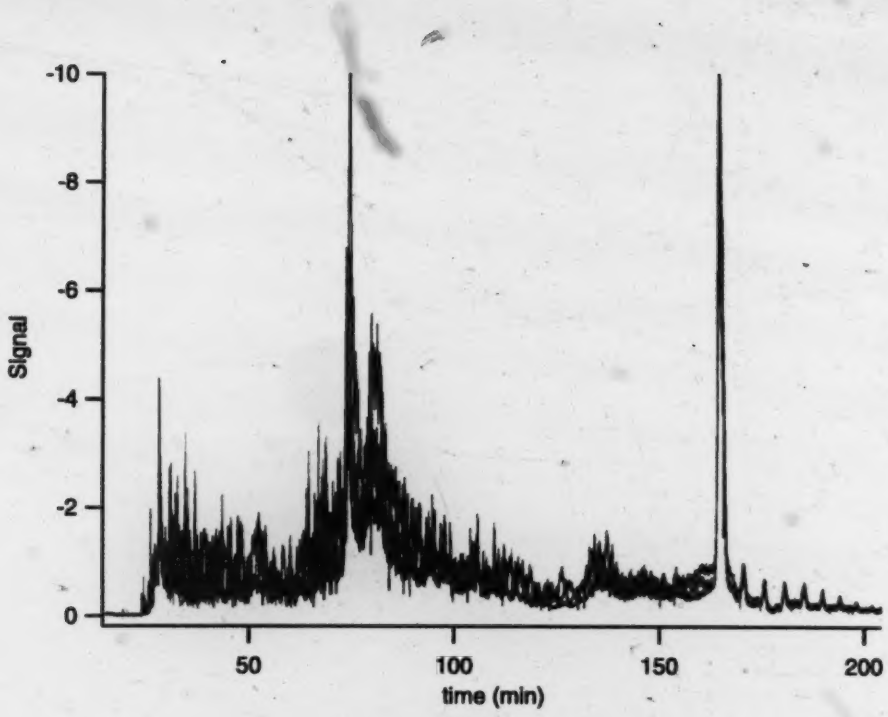


Figure 7.1. Sample Artifacts from PCR templates.

DNA sequencing run from wt M13 using a PCR product as template. Run on 6%T gel, at 150 V/cm and 40°C. At around 60 minutes, there is a small stretch. The remainder of the run contains lumps, jumps in the signal.

was not present in the ds plasmid preparations. The double-stranded plasmid templates were treated with RNA degrading (RNAse) enzymes before cycle sequencing reactions.

The evenly spaced peaks at the end of the run may be caused by of the presence of templates of various lengths, *i.e.* several PCR products. This artifact can be easily avoided by running an aliquot of the PCR reaction on a high concentration gel. If several DNA bands are present, indicating several PCR products, the individual bands can be cut from the gel and the separated DNA used as a sequencing template.

In sequencing wt M13, about 25% of the 640 electropherograms were ruined by the lump sample artifacts. Despite the ease in preparing the PCR templates, they were abandoned early in this wt M13 sequencing exercise for ds plasmid templates.

Staphylococcus Sequencing

It was later in sequencing *Staphylococcus* samples that the PCR products were used as sequencing templates. A fraction of the AB PCR product was run on an agarose gel to verify that only one PCR product was present. The PCR template was cleaned through a Qiagen PCR purification step. This step removed DNA fragments smaller than 100 bp, RNA, the A and B PCR primers, nucleotides and enzymes.

These above precautions reduced the occurrence of the catastrophic lumps observed in Figure 7.1. However, small lumps were still observed in the electropherograms but they did not appear consistently. Two different sample preparations from the same template of *S. saprophyticus* are shown in Figure 7.2. The lump artifact only appeared in the top electropherogram from data file: pr961106-2_cap2, 100 to 109 bases after the sequencing primer. The lump was not present in the bottom electropherogram (file: sb961119-1_cap2).

Stretch artifacts were also not eliminated from the sequencing run by cleaning the PCR template with the Qiagen kit. Figure 7.3, shows a stretch that occurred in two different cycle sequencing samples prepared from the same PCR template (files: pr961106-2_cap2 and sb961119-1_cap2). At 38°C the stretch was observed in both *S. saprophyticus* samples, 209 to 246 bases after the sequencing primer. The sequence was:

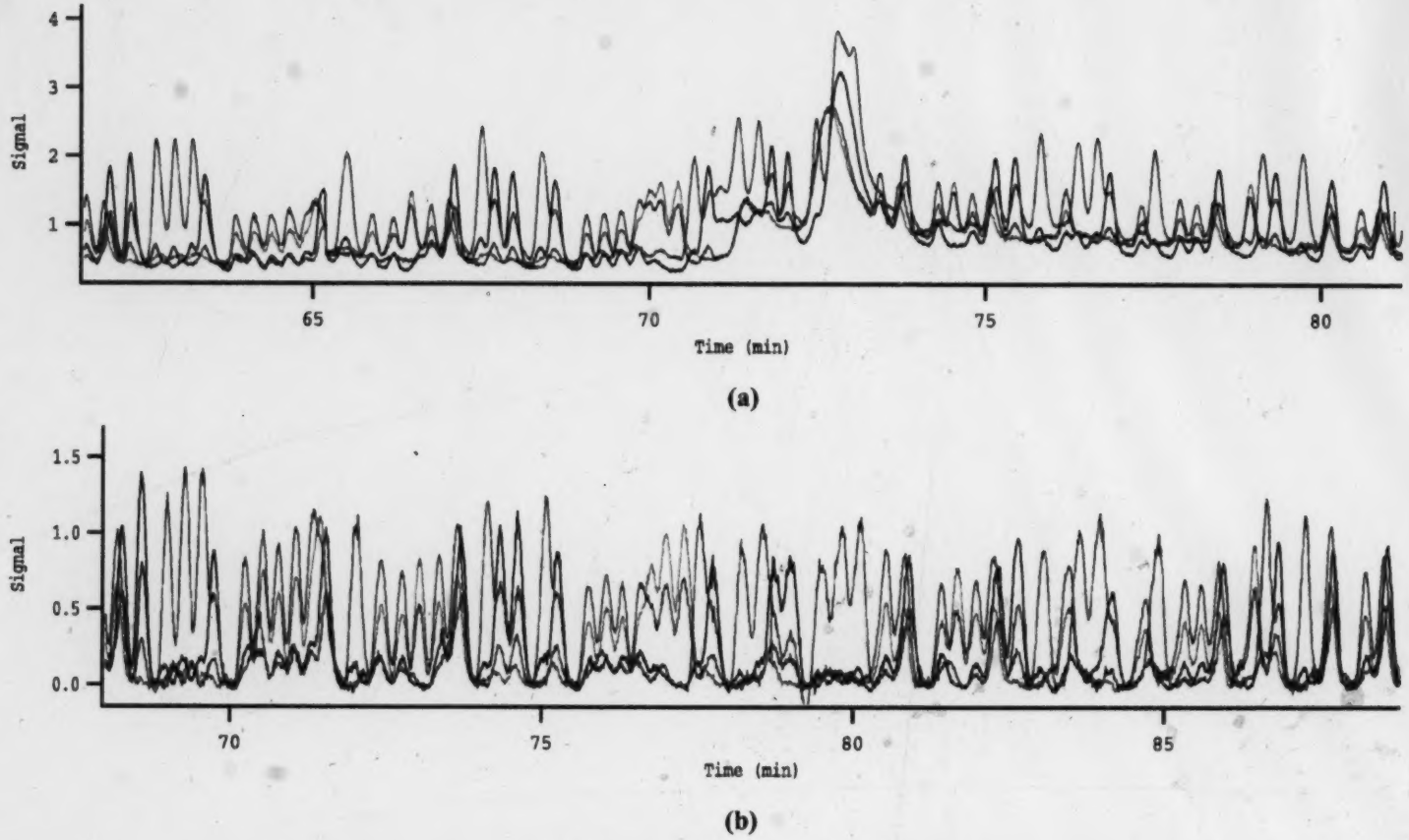
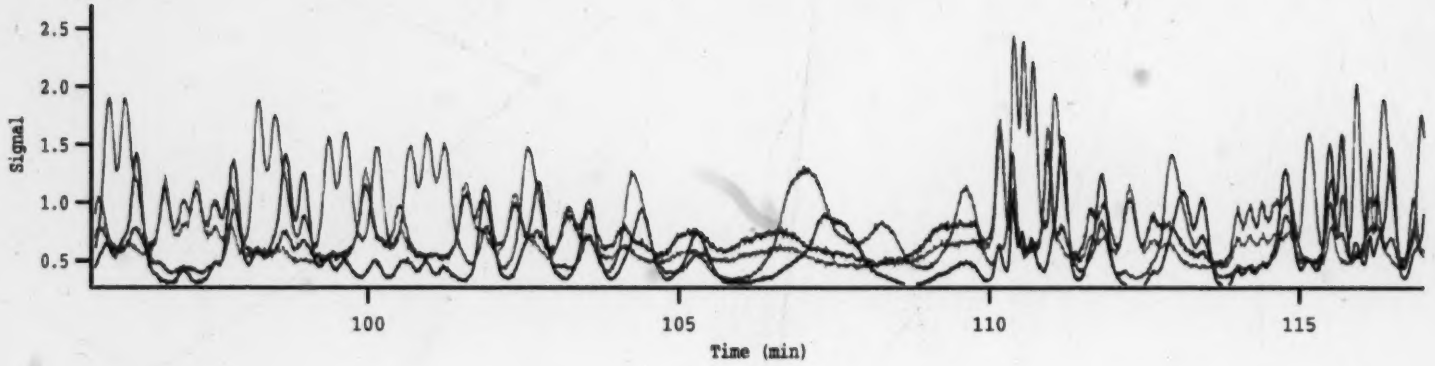
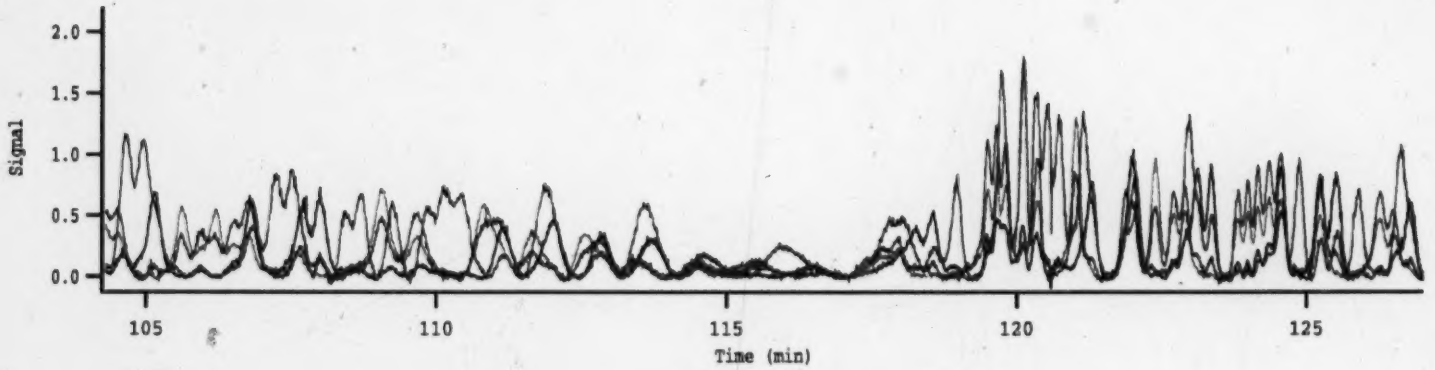


Figure 7.2. The Lump Artifact in Two Different Sample Preparations.

Two different sequencing sample preparations using the same PCR template of *S. saprophyticus*. Run at 38°C on the 5-capillary instrument and Plexiglas heater. In a), the sample exhibits the lump and in b), the lump is not present.



(a)



(b)

Figure 7.3. The Stretch Artifact in Two Different Sample Preparations.

Two different sequencing sample preparations using the same PCR template of *S. saprophyticus*. Run at 38°C on the 5-capillary instrument and Plexiglas capillary heater. In a), and b), the stretch is present.

TTTCAGCTGCAGATGAAGAAATTGGTAAATACATTTAT, with the underlined portion representing the sequence lost in both runs.

These stretches were seen for ds PCR and plasmid templates but not observed with ss templates. The stretch did not always occur in different sample preparations from the same template. They were not unique to the 16-capillary instrument and were seen on the 5-capillary instrument. The stretches were seen with the Plexiglas heater boxes on both instruments as shown in Chapter 2 (Figure 2.21).

Some sample artifacts could be removed by increasing the temperature with the Peltier heater. Figure 7.4 shows the effects of temperature on a stretch that occurred in a PCR template sample of *S. lugdunensis*. The stretch occurred at about 348 to 358 bases after the sequencing primer with the following sequence: TCCCATCCCCC. In Figure 7.4.a, the sample was run at 45°C (file: sb970509-1_cap12). When the same sample was run at 60°C, and the stretch was removed as shown in Figure 7.4.b (file: dr970516-1_cap13).

While the stretches and lumps did not always appear between 35 to 45°C, the sample artifacts consistently disappeared from runs heated to 55 to 60°C. This disappearance indicated the sample artifacts were related to secondary structure that melted with the increased heat. Because these stretches were not observed with single stranded templates, the stretch artifact may be the result of interstrand hybridization from the use of ds templates. It was suggested that the stretch was due to the unheated ends of capillary. The double stranded DNA is not completely melted apart as it migrates through the unheated end of the capillary. Because it has twice the charge as a single stranded DNA fragment of the same length, the double stranded DNA has a larger electrophoretic mobility and it will migrate more quickly. Once it reaches the heated portion of the capillary, the two strands are melted apart, and the DNA migrates at the mobility of the single stranded fragment [9]. Heating the entire length of the capillary, including the cuvette end, would test this hypothesis.

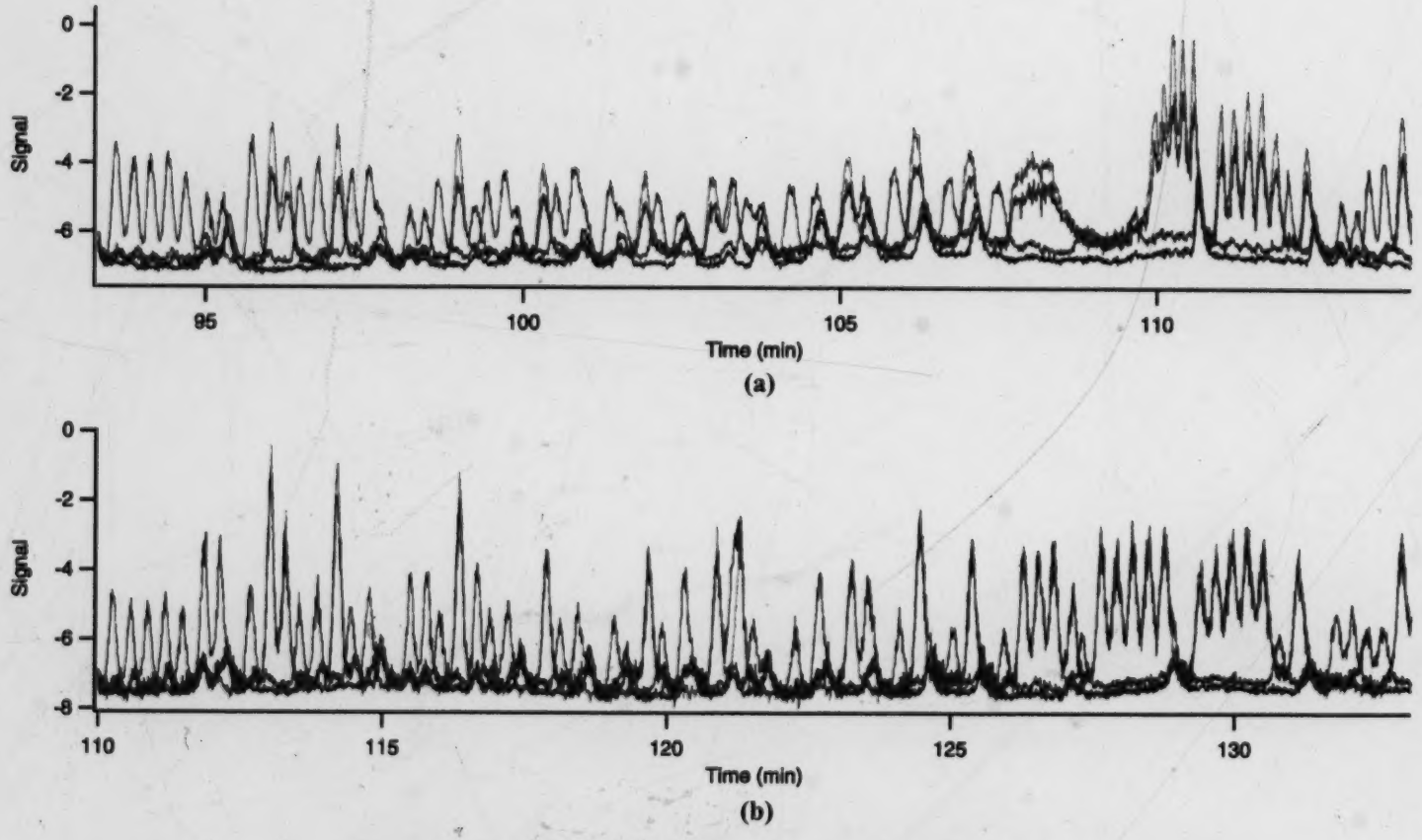


Figure 7.4. Melting the Stretch Artifact with the Peltier Heater.

DNA sequencing run of *S.lugdunensis* 345-358 bases after the primer and using a double-stranded PCR product as template. In a), the sample was run at 45°C and the stretch was observed. In b), the same sample was run at 60°C and the stretch was removed.

7.4. CONCLUSIONS

PCR templates, while easy to prepare, sometimes exhibited sample artifacts that destroyed sequencing runs. Lump sample artifacts were attributed to dirty template; they were removed with proper template treatment before sequencing. Stretch sample artifacts were caused by secondary structure between the two strands in the double stranded template. The stretches were eliminated from the sequencing run at temperature at 55 to 60°C.

Once the sample artifacts were eliminated, PCR templates were routinely used to prepare DNA sequencing templates quickly and produce satisfactory sequencing results.

7.5. REFERENCES

(1) Martin-Gallardo, A.; McCombie, W. R.; Gocayne, J. D.; Fitzgerald, M. G.; Wallace, S.; Lee, B. M. B.; Lamerdin, J.; Trapp, S.; Kelley, J. M.; Liu, L.-I.; Dubnick, M.; Johnston-Dow, L. A.; Kerlavage, A. R.; de Jong, P.; Carrano, A.; Fields, C.; Venter, J. C. *Nature Genetics* **1992,** *1*, 34-39.

(2) McCombie, W. R.; Martin-Gallardo, A. M. In *Automated DNA Sequencing and Analysis*; Adams, M. D., Fields, C., Venter, J. C., Eds.; Academic Press, Inc., **1994**, pp 159-166.

(3) Rowen, L.; Koop, B. F. In *Automated DNA Sequencing and Analysis*; Adams, M. D., Fields, C., Venter, J. C., Eds.; Academic Press, Inc., **1994**, pp 167-174.

(4) Pon, R. T.; Buck, G. A.; Niece, R. L.; Robertson, M.; Smith, A. J.; Spicer, E. *BioTechniques* **1994,** *17*, 526-534.

(5) Carothers, A. M.; Urlaub, G.; Mucha, J.; Grunberger, D.; Chasin, L. A. *BioTechniques* **1989,** *7*, 494-499.

(6) Du, Z.; Hood, L.; Wilson, R. K. *Methods in Enzymology* **1993,** *218*, 104-21.

(7) Hyder, S. M.; Hu, C.; Needleman, D. S.; Sonoda, Y.; Wang, X.-Y.; Baker, V. V. *BioTechniques* **1994,** *17*, 478-482.

(8) Hjerten, S. *Journal of Chromatography* **1985,** *347*, 181-198.

(9) Voss, K. O., personal communication, **1997**.

Chapter 8: Conclusions and Future Work

8.1. CONCLUSIONS

I have presented a 16-capillary DNA sequencer based on a micromachined sheath-flow cuvette for laser-induced fluorescence and avalanche photodiode detection. I reported a detection limit of 5×10^{-12} M fluorescein.

The separation instrument was modified to facilitate its use. The sample injection with a gold-plated microtitre plate was simple and a 16-capillary bomb refilled all the capillaries simultaneously. The compatibility of gas and heated capillaries for gel refilling will require further investigation. Heating the capillaries precisely to 55°C during the sequencing run removed sequencing artifacts that ruined the accuracy of sequencing.

The instrument successfully demonstrated four-color DNA sequencing on all 16 capillaries. The Heat Shock Protein-60 kDa (HSP60) gene from six *Staphylococcus* species were sequenced. These sequences were submitted to GenBank to create a database for a bacterial identification project. Perhaps this HSP60 PCR method can be used in combination with other PCR assays for genes unique to antibiotic resistant genes.

The 16-capillary sequencer had an average read length of about 500 bases per capillary at >97% accuracy acquired in about 3 h. Assuming 4-h turn around time for sample injection, sample run and gel refilling (20 min), at least 3-runs per day could be performed. Based on fifty, 5-day weeks per year, the predicted sequencing rate would be 8 kbases/run, 24 kbases/day, 120 kbases/week, and 6 Mbases/year.

The factor limiting capillary DNA sequencing was the gels. In order to be a truly high-throughput sequencer, the gels must perform reliably. In the *Staphylococcus* sequencing, about one-fifth of the capillaries generated useful data. An improved gel matrix must be developed that has a low viscosity and resolves long stretches of DNA sequence. The gel problem may eventually be solved by sequencing without gels (end-labeled free solution electrophoresis, ELFSE) [1].

The 16-capillary instrument played an important role in the evolution of multiple-capillary instruments. Two significant design changes are required for future generations of multiple-capillary instruments. First, the filter wheel with four bandpass filters are not efficient for larger-capillary instruments. A low duty-cycle arose from collecting the

fluorescent signal in each channel sequentially as the bandpass filter rotated into position. A dispersion element such as a diffraction grating or prism enables detection of the fluorescence spectra of the four dyes simultaneously. Second, point detectors like the GRIN lenses and avalanche photodiodes were an inexpensive but impractical to optically align for more than 16 capillaries. The fluorescence images are more efficiently detected by charge-coupled device array (CCD) cameras with their large photosensitive areas. Combined with the color dispersion, the CCD cameras will have increased duty cycles and higher collection efficiencies.

This group has now also built a 32-capillary sequencer, based on a micromachined cuvette, camera lenses, and four-color dispersion onto a CCD detector [2]. The linear array of capillaries, however, will soon be replaced by 2-dimensional arrays of capillaries such as the 8×12 array of the 96-capillary instrument [3].

The future of DNA sequencing instruments may lie in the microchip devices in a miniaturized total analysis systems [4]. These devices are essentially capillaries, microchannels, etched on the surface of planar glass substrates. Mathies's group has microfabricated 12 capillary array electrophoresis chip [5]. Three problems are apparent. The detection optics have not concurrently miniaturized with the microchip devices. The sample loading is still reminiscent of slab gel wells and will not be simple to automate as for capillaries. The microchannel length may approach lengths of normal capillaries in order to achieve reasonable sequencing resolutions. Lawrence Livermore National Laboratory has built a hybrid slab gel sequencer where long channels are etched along the length of a slab gel plate [6].

DNA sequencing fragments have been detected by matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry [7] and electrospray ionization with ion cyclotron resonance mass spectrometry [8]. An interesting approach is sequencing-by-hybridization [9, 10]. The target DNA is fluorescently labeled and will bind to its complement which is covalently bound onto a chip. The binding will lead to a two-dimensional picture that is translated into DNA sequence.

In 1996, the US Human Genome Project decided to focus less on mapping and pushed to sequence DNA using the current technology available, a multitude of PE

Applied Biosystems slab gel sequencers [11, 12]. This "brute-force" method would ensure that the genome sequence would be delivered into the hands of biologists by 2005. This attitude was understandable as the new technology to speed up sequencing and reduce costs was taking longer to develop than anticipated.

However there are many organisms other than human to sequence such as the mouse, the fruit fly and agriculturally significant plants such as corn and rice. In January 1997, there were an estimated 30 genome projects ongoing in the world [13]. In 1993, the Department of Energy Subsurface Culture Collection initiated an effort to phylogenetically characterize the 10 000 strains of bacteria in their collection [14]. Before this project, essentially the origin and the morphology was all that was known about the bacteria in the collection. Only 4% of the strains have been analyzed by 16S rRNA gene sequences to date.

The initial development of technology to sequence DNA from radioactive probes to fluorescent instruments has revolutionized biology and medicine by allowing it to be feasible for most researcher to study organisms' genes [15, 16]. The next step is to develop instrumentation that will allow studying genomes to be similarly feasible.

8.2. REFERENCES

(1) Slater, G. W.; Desruisseaux, C.; Villeneuve, C.; Guo, H. L.; Drouin, G. *Electrophoresis* **1995,** *16*, 704-12.

(2) Jiang, R.; Lewis, D. F.; Dovichi, N. J., in preparation.

(3) Zhang, J. Z.; Dovichi, N. J., in preparation.

(4) Kopp, M. U.; Crabtree, H. J.; Manz, A. *Current Opinion in Chemical Biology* **1997,** *1*, 410-419.

(5) Woolley, A. T.; Sensabaugh, G. F.; Mathies, R. A. *Analytical Chemistry* **1997,** *69*, 2181-86.

(6) Brewer, L. B.; Davidson, J. C.; Balch, J. W.; Carrano, A. V. *Electrophoresis* **1995,** *16*, 1846-50.

(7) Taranenko, N. I.; Chung, C. N.; Zhu, Y. F.; Allman, S. L.; Golovlev, V.; Isola, N. R.; Martin, S. A.; Haff, L. A.; Chen, C. H. *Rapid Communications in Mass Spectrometry* **1997,** *11*, 386-92.

(8) Smith, R. D.; Cheng, X.; Hofstadler, S. A.; Bruce, J. A.; Edmonds, C. G., *DOE Human Genome Program Contractor-Grantee Workshop IV*, Sante Fe, NM **1995**.

(9) Pease, A. C.; Solas, D.; Sullivan, E. J.; Cronin, M. T.; Holmes, C. P.; Fodor, S. P. *Proceedings of the National Academy of Sciences of the United States of America* **1994,** *91*, 5022-6.

(10) Chee, M.; Yang, R.; Hubbell, E.; Berno, A.; Huang, X. C.; Stern, D.; Winkler, J.; Lockhart, D. J.; Morris, M. S.; Fodor, S. P. *Science* **1996,** *274*, 610-14.

(11) Marshall, E. *Science* **1995,** *268*, 1270-1271.

(12) Venter, J. C.; Smith, H. O.; Hood, L. *Nature* **1996,** *381*, 364-366.

(13) Clayton, R. A.; White, O.; Ketchum, K. A.; Venter, J. C. *Nature* **1997,** *387*, 459-462.

(14) Balkwill, D. L.; Reeves, R. H.; Drake, G. R.; Reeves, J. Y.; Crocker, F. H.; Baldwin King, M.; Boone, D. R. *FEMS Microbiological Reviews* **1997,** *20*, 201-216.

(15) Olsen, G. J.; Woese, C. R.; Overbeek, R. *Journal of Bacteriology* **1994,** *176*, 1-6.

(16) Fraser, C. M.; Fleischmann, R. D. *Electrophoresis* **1997,** *18*, 1207-16.

Appendix I: DNA Sequencing Reagents

Amersham Life Science

Thermo Sequenase fluorescently labeled primer cycle sequencing kit RPN 2436

A reagent: Tris-HCl (pH=9.5), $MgCl_2$, Tween™ 20, Nonidet™ P-40, 2-mercaptoethanol, dATP, dCTP, dGTP, dTTP, ddATP, thermostable pyrophosphatase, and Thermo Sequenase DNA polymerase.

C reagent: Tris-HCl (pH=9.5), $MgCl_2$, Tween™ 20, Nonidet™ P-40, 2-mercaptoethanol, dATP, dCTP, dGTP, dTTP, ddCTP, thermostable pyrophosphatase, and Thermo Sequenase DNA polymerase.

G reagent: Tris-HCl (pH=9.5), $MgCl_2$, Tween™ 20, Nonidet™ P-40, 2-mercaptoethanol, dATP, dCTP, dGTP, dTTP, ddGTP, thermostable pyrophosphatase, and Thermo Sequenase DNA polymerase.

T reagent: Tris-HCl (pH=9.5), $MgCl_2$, Tween™ 20, Nonidet™ P-40, 2-mercaptoethanol, dATP, dCTP, dGTP, dTTP, ddTTP, thermostable pyrophosphatase, and Thermo Sequenase DNA polymerase.

Appendix II: Partial HSP60 Gene Sequences

S.aureus

Partial HSP60 Gene Sequence

```
1     TGCAACAGTA TTAGCTCAAG CAATGATTCA AGAAGGCTTG AAAAATGTTA
      ACGTTGTCAT AATCGAGTTC GTTACTAAGT TCTTCCGAAC TTTTTACAAT
51    CAAGTGGTGC GAACCCAGTT GGTTTACGAC AAGGTATCGA CAAAGCAGTT
      GTTCACCACG CTTGGGTCAA CCAAATGCTG TTCCATAGCT GTTTCGTCAA
101   AAAGTTGCTG TTGAAGCGTT ACATGAAAAT TCTCAAAAAG TTGAAAATAA
      TTTCAACGAC AACTTCGCAA TGTACTTTTA AGAGTTTTTC AACTTTTATT
151   AAATGAAATT GCGCAAGTAG GTGCGATTTC AGCAGCAGAT GAAGAAATTG
      TTTACTTTAA CGCGTTCATC CACGCTAAAG TCGTCGTCTA CTTCTTTAAC
201   GACGTTATAT TTCTGAAGCT ATGGAAAAAG TAGGTAACGA TGGTGTCATT
      CTGCAATATA AAGACTTCGA TACCTTTTTC ATCCATTGCT ACCACAGTAA
251   ACAATTGAAG AATCAAATGG ACTAAACACT GAACTAGAAG TGGTTGAAGG
      TGTTAACTTC TTAGTTTACC TGATTTGTGA CTTGATCTTC ACCAACTTCC
301   TATGCAATTT GATCGTGGTT ATCAATCACC GTATATGGTT ACTGATTCAG
      ATACGTTAAA CTAGCACCAA TAGTTAGTGG CATATACCAA TGACTAAGTC
351   ATAAAATGGT TGCTGAATTA GAACGCCCAT ACATTTTAGT AACAGATAAG
      TATTTTACCA ACGACTTAAT CTTGCGGGTA TGTAAAATCA TTGTCTATTC
401   AAAATCTCGT CTTTCCAAGA TATCTTACCT TTATTAGAAC AAGTGGTTCA
      TTTTAGAGCA GAAAGGTTCT ATAGAATGGA AATAATCTTG TTCACCAAGT
451   ATCTAATCGT CCAATCTTAA TTGTAGCTGA TGAAGTTGAA GGCGATGCAT
      TAGATTAGCA GGTTAGAATT AACATCGACT ACTTCAACTT CCGCTACGTA
501   TAACAAATAT CGTGCTAAAC CGTATGCGTG GCACATTTAC AGCTGTTGCA
      ATTGTTTATA GCACGATTTG GCATACGCAC CGTGTAAATG TCGACAACGT
551   GTA
      CAT
```

S.epidermidis
Partial HSP60 Gene Sequence

```
1    AGCAACAGTT TTAGCACAAT CAATGATTCA GGAAGGTCTT AAGAATGTTA
     TCGTTGTCAA AATCGTGTTA GTTACTAAGT CCTTCCAGAA TTCTTACAAT
51   CAAGTGGTGC AAATCCTGTA GGCTTAAGAC AAGGTATTGA CAAAGCAGTG
     GTTCACCACG TTTAGGACAT CCGAATTCTG TTCCATAACT GTTTCGTCAC
101  CAAGTGGCTA TAGAAGCGCT CCATGAGATT TCTCAAAAGG TTGAAAATAA
     GTTCACCGAT ATCTTCGCGA GGTACTCTAA AGAGTTTTCC AACTTTTATT
151  GAACGAGATA GCGCAAGTTG GAGCTATTTC AGCAGCAGAT GAAGAAATCG
     CTTGCTCTAT CGCGTTCAAC CTCGATAAAG TCGTCGTCTA CTTCTTTAGC
201  GTCGCTACAT TTCTGAAGCA ATGGATAAAG TAGGTAACGA TGGCGTTATC
     CAGCGATGTA AAGACTTCGT TACCTATTTC ATCCATTGCT ACCGCAATAG
251  ACTATTGAAG AATCAAATGG GTTTAATACA GAATTAGAAG TAGTTGAAGG
     TGATAACTTC TTAGTTTACC CAAATTATGT CTTAATCTTC ATCAACTTCC
301  AATGCAATTT GATCGCGGTT ATCAATCACC ATATATGGTA ACTGACTCAG
     TTACGTTAAA CTAGCGCCAA TAGTTAGTGG TATATACCAT TGACTGAGTC
351  ATAAAATGAT AGCTGAATTA GAACGTCCAT ATATATTAGT AACGGATAAG
     TATTTTACTA TCGACTTAAT CTTGCAGGTA TATATAATCA TTGCCTATTC
401  AAAATTTCAT CATTCCAAGA TATTCTTCCA TTATTAGAAC AAGTTGTGCA
     TTTTAAAGTA GTAAGGTTCT ATAAGAAGGT AATAATCTTG TTCAACACGT
451  GGCTAGTCGA CCAATTTTAA TTGTTGCGGA TGAAGTAGAA GGCGATGCAC
     CCGATCAGCT GGTTAAAATT AACAACGCCT ACTTCATCTT CCGCTACGTG
501  TTACTAATAT TGTTTTAAAC CGTATGCGTG GAACATTTAC TGCTGTAGCA
     AATGATTATA ACAAAATTTG GCATACGCAC CTTGTAAATG ACGACATCGT
551  GTC
     CAG
```

S.haemolyticus
Partial HSP60 Gene Sequence

```
1     TGCAACTGTA TTAGCACAAG CAATGATTCA AGAAGGATTG AAGAACGTAA
      ACGTTGACAT AATCGTGTTC GTTACTAAGT TCTTCCTAAC TTCTTGCATT
51    CAAGTGGTGC TAACCCAGTA GGATTACGTG AGGGTATTGA TAAAGCAGTA
      GTTCACCACG ATTGGGTCAT CCTAATGCAC TCCCATAACT ATTTCGTCAT
101   AGAGTAGCTG TACAAGCTCT ACACGATATT TCTCAAAAAG TTGAAAATAA
      TCTCATCGAC ATGTTCGAGA TGTGCTATAA AGAGTTTTTC AACTTTTATT
151   AAATGAAATC GCACAAGTAG GTGCAATTTC TGCAGCAGAT GAAGAAATTG
      TTTACTTTAG CGTGTTCATC CACGTTAAAG ACGTCGTCTA CTTCTTTAAC
201   GTAAATACAT TTCTGAAGCA ATGGATAAAG TTGGTAATGA TGGTGTTATT
      CATTTATGTA AAGACTTCGT TACCTATTTC AACCATTACT ACCACAATAA
251   ACCATTGAAG AATCAAATGG ACTAGATACA GAATTAGAAG TTGTTGAAGG
      TGGTAACTTC TTAGTTTACC TGATCTATGT CTTAATCTTC AACAACTTCC
301   TATGCAATTT GATAGAGGGT ATCAATCACC ATACATGGTT ACAGATTCAG
      ATACGTTAAA CTATCTCCCA TAGTTAGTGG TATGTACCAA TGTCTAAGTC
351   ATAAAATGAT TGCTGAGTTA GAAAGACCAT ATATTTTAGT AACAGATAAA
      TATTTTACTA ACGACTCAAT CTTTCTGGTA TATAAAATCA TTGTCTATTT
401   AAAATCTCTT CATTCCAAGA TATCTTACCT TTATTAGAAC AAGTTGTTCA
      TTTTAGAGAA GTAAGGTTCT ATAGAATGGA AATAATCTTG TTCAACAAGT
451   ATCTAGTCGT CCAATCTTAA TTGTAGCAGA CGAAGTTGAA GGAGACGCTT
      TAGATCAGCA GGTTAGAATT AACATCGTCT GCTTCAACTT CCTCTGCGAA
501   TAACAAATAT CGTATTAAAC CGTATGCGTG GAACATTTAC TGCAGTTGCA
      ATTGTTTATA GCATAATTTG GCATACGCAC CTTGTAAATG ACGTCAACGT
551   GTA
      CAT
```

S.lugdunesis
Partial HSP60 Gene Sequence

```
1     TGCTACAGTG CTAGCACAAG CAATGATTCA AGAAGGACTT AAAAACGTAA
      ACGATGTCAC GATCGTGTTC GTTACTAAGT TCTTCCTGAA TTTTTGCATT
51    CGAGTGGTGC AAATCCAGTT GGCTTACGTC AAGGTATTGA TAAAGCAGTT
      GCTCACCACG TTTAGGTCAA CCGAATGCAG TTCCATAACT ATTTCGTCAA
101   AAAGTAGCTA TCGAAGCATT ACATGATATA TCACAAAAAG TTGAAAATAA
      TTTCATCGAT AGCTTCGTAA TGTACTATAT AGTGTTTTTC AACTTTTATT
151   AAATGAAATT GCACAAGTTG GGGCTATTTC TGCAGCAGAT GAAGAGATTG
      TTTACTTTAA CGTGTTCAAC CCCGATAAAG ACGTCGTCTA CTTCTCTAAC
201   GACAATATAT TTCAGAAGCT ATGGATAAAG TAGGTAATGA TGGCGTAATA
      CTGTTATATA AAGTCTTCGA TACCTATTTC ATCCATTACT ACCGCATTAT
251   ACTATTGAAG AATCTAGTGG TTTCAATACT GAACTTGAAG TTGTAGAAGG
      TGATAACTTC TTAGATCACC AAAGTTATGA CTTGAACTTC AACATCTTCC
301   TATGCAATTT GATAGAGGTT ACCAATCACC ATATATGGTT ACGGATTCAG
      ATACGTTAAA CTATCTCCAA TGGTTAGTGG TATATACCAA TGCCTAAGTC
351   ATAAAATGAC TGCAGAATTA GAACGTCCAT ATATATTAGT TACAGATAAA
      TATTTTACTG ACGTCTTAAT CTTGCAGGTA TATATAATCA ATGTCTATTT
401   AAAATATCTT CATTCCAAGA TATTTTACCT TTACTAGAAC AAGTCGTTCA
      TTTTATAGAA GTAAGGTTCT ATAAAATGGA AATGATCTTG TTCAGCAAGT
451   ATCTAGTCGA CCAATATTAA TTGTTGCTGA TGAAGTAGAA GGTGATGCAT
      TAGATCAGCT GGTTATAATT AACAACGACT ACTTCATCTT CCACTACGTA
501   TAACTAATAT CGTTCTAAAC CGCATGCGTG GAACATTCAC TGCAGTTGCT
      ATTGATTATA GCAAGATTTG GCGTACGCAC CTTGTAAGTG ACGTCAACGA
551   GTT
      CAA
```

S.saprophyticus
Partial HSP60 Gene Sequence

```
1    AGCAACAGTG CTAGCACAAG CTATGATTCA AGAAGGTTTG AAAAACGTAA
     TCGTTGTCAC GATCGTGTTC GATACTAAGT TCTTCCAAAC TTTTTGCATT
51   CAAGTGGTGA AAACCCTGTT GGTTTAAGAC AAGGTATTGA TAAAGCTGTA
     GTTCACCACT TTTGGGACAA CCAAATTCTG TTCCATAACT ATTTCGACAT
101  GAAGTAGCAA TTGAGGCATT ACATGAAATT TCACAAAATG TAGATAATAA
     CTTCATCGTT AACTCCGTAA TGTACTTTAA AGTGTTTTAC ATCTATTATT
151  AAATGAAATT GCACAAGTTG GTTCTATTTC AGCTGCAGAT GAAGAAATTG
     TTTACTTTAA CGTGTTCAAC CAAGATAAAG TCGACGTCTA CTTCTTTAAC
201  GTAAATACAT TTCTGAAGCA ATGGAAAAAG TAGGTAATGA TGGTGTCATC
     CATTTATGTA AAGACTTCGT TACCTTTTTC ATCCATTACT ACCACAGTAG
251  ACTATTGAAG AGTCTAGTGG TTTTAATACA GAATTAGAAG TAGTAGAAGG
     TGATAACTTC TCAGATCACC AAAATTATGT CTTAATCTTC ATCATCTTCC
301  TATGCAATTT GACCGAGGTT ATCAATCACC ATATATGGTG ACTGACTCAG
     ATACGTTAAA CTGGCTCCAA TAGTTAGTGG TATATACCAC TGACTGAGTC
351  ATAAAATGGT CGCTGATCTT GAAAGACCAT ATATTTTAAT TACTGATAAA
     TATTTTACCA GCGACTAGAA CTTTCTGGTA TATAAAATTA ATGACTATTT
401  AAGATTTCAT CATTCCAAGA CATCTTACCG TTATTAGAAC AAGTCGTACA
     TTCTAAAGTA GTAAGGTTCT GTAGAATGGC AATAATCTTG TTCAGCATGT
451  ATCAAATCGT CCTATCTTAA TTGTTGCTGA TGACGTTGAA GGCGATGCAT
     TAGTTTAGCA GGATAGAATT AACAACGACT ACTGCAACTT CCGCTACGTA
501  TAACTAATAT CGTGCTTAAC CGTATGCGTG GCACATTTAC TGCTGTTGCA
     ATTGATTATA GCACGAATTG GCATACGCAC CGTGTAAATG ACGACAACGT
551  GTG
     CAC
```

S.schleiferi
Partial HSP60 Gene Sequence

```
1     AGCGACTGTT TTAGCACAAG CTATGATTCA AGAAGGACTT AAAAATGTAA
      TCGCTGACAA AATCGTGTTC GATACTAAGT TCTTCCTGAA TTTTTACATT
51    CAAGTGGCGC AAACCCAGTA GGTATCCGTC AAGGTATTGA TAAAGCGGTA
      GTTCACCGCG TTTGGGTCAT CCATAGGCAG TTCCATAACT ATTTCGCCAT
101   GCTGTAGCGA TTGAAGCACT TCATAATATT TCTCAAAAAG TAGAAAATAA
      CGACATCGCT AACTTCGTGA AGTATTATAA AGAGTTTTTC ATCTTTTATT
151   AGAAGAAATC GCGCAAGTAG GTGCTATTTC TGCAGCTGAT GAAGAAGTAG
      TCTTCTTTAG CGCGTTCATC CACGATAAAG ACGTCGACTA CTTCTTCATC
201   GTCGTTATAT TTCTGAAGCT ATGGAAAAAG TAGGTAATGA CGGTGTTATC
      CAGCAATATA AAGACTTCGA TACCTTTTTC ATCCATTACT GCCACAATAG
251   TCTATTGAAG AATCAAGCGG TTTTAATACT GAATTAGATG TAGTGGAAGG
      AGATAACTTC TTAGTTCGCC AAAATTATGA CTTAATCTAC ATCACCTTCC
301   TATGCAATTC GATCGTGGAT ACCAATCTCC ATATATGGTT ACTGATTCGG
      ATACGTTAAG CTAGCACCTA TGGTTAGAGG TATATACCAA TGACTAAGCC
351   ATAAAATGAC TGCTGAGTTA GAAAAACCTT ATATTTTAAT CACGGATAAA
      TATTTTACTG ACGACTCAAT CTTTTTGGAA TATAAAATTA GTGCCTATTT
401   AAAATCTCAT CTTTCCAAGA TATCCTACCT TTATTAGAGC AAATCGTTCA
      TTTTAGAGTA GAAAGGTTCT ATAGGATGGA AATAATCTCG TTTAGCAAGT
451   ATCTAATCGC CCAATTTTAA TTGTTGCTGA TGAAGTTGAA GGCGACGCTT
      TAGATTAGCG GGTTAAAATT AACAACGACT ACTTCAACTT CCGCTGCGAA
501   TAACAAACTT AGTATTGAAC CGTATGCGTG GTACATTTAC TGCCGTTGCA
      ATTGTTTGAA TCATAACTTG GCATACGCAC CATGTAAATG ACGGCAACGT
551   GTA
      CAT
```